

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 02:53:24 ; Search time 327 Seconds  
(without alignments)  
4977.863 Million cell updates/sec

Title: US-09-744-315-7

Perfect score: 603  
Sequence: 1 caggaacagcctctcctgc.....aaaagataagtcgctgc 603

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: N.GeneSeq\_19Jun03.\*

1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT:\*  
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT:\*  
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT:\*  
25: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	603	100.0	603	AAZ50577	Human epidermal pr
2	511.4	84.8	666	AAK51648	Human polynucleoti
3	509.8	84.5	617	AAK52632	Human polynucleoti
4	355.4	58.9	486	21 AACT7838	Human cancer assoc
5	142.6	23.6	689	24 ABT09264	Phase-1 Rat CT gen
6	134.4	22.3	148	24 ABRK5393	Human head and nec
7	104.8	17.4	422	22 AAK52162	Human polynucleoti
8	104.8	17.4	422	22 AAK53146	Human polynucleoti

9	95.6	15.9	540	23 AAZ56924	DNA encoding novel
10	93.8	15.6	414	22 AAK52164	Human polynucleoti
11	93.8	15.6	414	22 AAK53148	Human polynucleoti
12	85.8	14.2	635	22 AAK81731	Human secreted pro
13	72.4	12.0	526	22 AAK85555	CDNA encoding CDF
14	71	11.8	528	21 AAZ50579	Human epidermal pr
15	68.8	11.4	678	24 ABO61174	Human epidermal pr
16	62.4	10.3	697	21 AAZ50578	Human epidermal pr
17	44.6	7.4	403	21 ABR10240	Trinucleotide repe
18	43.4	7.2	4335	23 AAZ79854	DNA encoding novel
19	43.2	7.2	397	20 AAK89891	Spinocezebellar at
20	43.2	7.2	1159	21 AAK59240	An Ecocri fragment
21	42.6	7.1	844	24 ABR9517	Gene #2215 used to
22	42.4	7.0	515	20 AAK18930	Human huntingtin g
23	42.4	7.0	600	25 ABR28178	Huntington's disea
24	42.4	7.0	614	25 ABR281762	Huntington's disea
25	42.4	7.0	10348	22 AAZ23426	Human huntingtin p
26	42.4	7.0	10348	22 ABR06651	Human huntingtin p
27	42.4	7.0	10348	23 ABR51971	DNA encoding novel
28	42.4	7.0	10366	15 AAK67401	Huntingtin DVA\ITI
29	42.4	7.0	10366	18 AAT97924	Composite sequence
30	42.4	7.0	10366	19 AAV05828	Human huntingtin c
31	42.2	7.0	477	21 AAZ44307	Human SCA7 genomic
32	42	7.0	1551	24 ABR19450	Mouse ischaemic co
33	42	7.0	1551	25 ABR76651	Gene #1748 used to
34	41.8	6.9	2614	24 ABR36100	CDNA sequence #491
35	41.8	6.9	6604	24 ABR95250	Gene #1748 used to
36	41.8	6.9	6788	24 ABRK3506	Human CDNA differe
37	41.8	6.9	6794	21 AAZ45597	CDNA sequence of h
38	41.6	6.9	336	20 AAZ00010	Antifreeze peptide
39	41.6	6.9	336	20 AAX8092	Synthetic antifree
40	41.6	6.9	657	24 AAD45051	Human zsl963 degen
41	41.6	6.9	657	24 ABR52634	Human secreted sal
42	41.6	6.9	657	24 AAZ20592	Human zsl963 degen
43	41.6	6.9	657	25 ABR35395	Human zsl963 degen
44	41.6	6.9	1037	21 AAK59241	Exons B, C and A o
45	41.6	6.9	1472	21 AAZ59241	Exons D, C, B and

## ALIGNMENTS

RESULT 1	AAZ50577	standard; cDNA; 603 BP.
ID	AAZ50577	
AC	AAZ50577;	
XX	23-MAY-2000	(first entry)
XX	Human epidermal protein-1 cDNA.	
XX	Human epidermal protein-1; HBP1; epithelial disorder; scabies;	
XX	dyshidrotic eczema; cell proliferative disorder; actinic keratosis;	
XX	arteriosclerosis; autoimmune disorder; inflammatory disorder;	
XX	acquired immune deficiency syndrome; AIDS; Addison's disease; antiHIV;	
XX	dematological; antitartaroclerotic; antiinflammatory;	
XX	immunosuppressive; ss.	
XX	Homo sapiens.	
XX	OS	
XX	Key	Location/Qualifiers
XX	FT	56..334
XX	FT	/*tag= a
XX	FT	/product= "Human epidermal protein-1"
XX	XX	WO200006727-A2.
XX	PD	10-FEB-2000.
XX	XX	27-JUL-1999;
XX	XX	99MO-US17107.
XX	XX	28-JUL-1998;
XX	XX	98US-0155203.



Qy	1	CAGGAACAGCCTTCTCCTGCTCCTTGACCTGGACCAATCAATCTCCGCAAGATGTC	60
Db	43	CAGGAACAGCCTTCTCCTGCTCCTTGACCTGGACCAATCAATCTCCGCAAGATGTC	102
Qy	61	CTGCTCAGCAAGACCAAGAGCAATGCCAACCCCAACCAAGTGTCCCTCAACCAATGTCC	120
Db	103	CTGCTCAGCAAGACCAAGAGCAATGCCAACCCCAACCAAGTGTCCCTCAACCAATGTCC	167
Qy	121	CCCAAGAGCCAGTACAGTGTGTGCTCAGCTTCTCTGCTGTGCCCCAAGCTGTG	180
Db	163	CCCAAGAGCCAGTACAGTGTGTGCTCAGCTTCTCTGCTGTGCCCCAAGCTGTG	222
Qy	181	GGTCTGTGGCCCTAAGCTCCGAGGGGGGGCTTCTTGAAACAACAAGGGCCACACCG	240
Db	223	GGGCTGTGGCCCTAAGCTCCGAGGGGGGGCTTCTTGAAACAACAAGGGCCACACCG	287
Qy	241	ATGCGCGCGCCAGAGGCCCAATCTCTGTACAAGGGCAATGTGTACAGAGCGGGGGCTC	300
Db	283	ATGCGCGCGCCAGAGGCCCAATCTCTGTACAAGGGCAATGTGTACAGAGCGGGGGCTC	347
Qy	301	TGGGTGTGTGCAAGGTCTGGGGGGGCTGCTGATATCAATCTGATGCTGACACACG	360
Db	343	TGGGTGTGTGCAAGGTCTGGGGGGGCTGCTGATATCAATCTGATGCTGACACACG	407
Qy	361	ATCTTTGAGAGAAACAAGATTCCAAGAGGCCAAGAACGCCCATGTGACGATGCTT	420
Db	403	AT-TTTGAGAGAAACAAGATTCCAAGAGGCCAAGAACGCCCATGTGATGATGATTT	461
Qy	421	CCCATATACCTCTCTTGACTTCAAGAGGTAGCTGAGAGTTTCTCTGTGGGGATCTG	480
Db	462	CCCATATACCTCTCTCTTGACTTCAAGAGGTAGCTGAGAGTTTCTCTGTGGAGATCTG	527
Qy	481	AGCTCTCCCAAGAGCACTTCTTTTATGTACAGATGTCAATGTCCCCCTAACCC	544
Db	522	AGCTCTCCCAAGAGCACTTCTTTTATGTACAGATGTCAATGTCCCCCTAACCC	583
Qy	541	TGTACTCGCAAGATTTGGCAGTGCTTGTGCCCAACTGTAAAAAGATAAGTTTC	597
Db	582	TGTACTCGTAAGATTTGGCAGTGCTTGTGCCCAACTGTAAAAATTAAGCTTC	638
RESULT 3			
AAKS2632			
ID	AAKS2632 standard; cDNA; 617 BP.		
XX			
AC	AAKS2632;		
XX			
DT	06-NOV-2001 (first entry)		
XX			
DE	Human polynucleotide SEQ ID NO 2161.		
XX			
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;		
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;		
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;		
KW	nervous system disorder; arthritis; inflammation; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200157190-A2.		
XX			
PD	09-AUG-2001.		
XX			
PF	05-FEB-2001; 2001MO-US04098.		
XX			
PR	03-FEB-2000; 2000US-0496914.		
PR	27-APR-2000; 2000US-0560875.		
PR	20-JUN-2000; 2000US-0598075.		
PR	19-JUL-2000; 2000US-0620325.		
PR	01-SEP-2000; 2000US-0654936.		
PR	15-SEP-2000; 2000US-0683261.		
PR	20-OCT-2000; 2000US-0683261.		
PR	30-NOV-2000; 2000US-0728422.		

XX	(HYSE-) HYSEQ INC.
PA	
XX	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
XX	Tang QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
P1	Zhuo Q, Yang T, Wejhrman T, Goodrich R;
P1	Xue M, Yang T, Wejhrman T, Goodrich R;
XX	WPI; 2001:476283/51.
DR	P-PSDB; AAM79499.
XX	
PT	Nucleic acids encoding polypeptides with cytokine-like activities,
PT	useful in diagnosis and gene therapy -
XX	
XX	Claim 1, Page 4528-4529; 622IP; English.
P8	
XX	The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC	encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC	cytokine, cell proliferation or cell differentiation or which may induce
CC	production of other cytokines in other cell populations. The
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC	peptide therapy. The polypeptides have various cytokine-like activities,
CC	e.g. stem cell growth factor activity, haematopoiesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	activin/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC	inflammation.
CC	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC	(AAM80020) are omitted as the relevant pages from the sequence listing
CC	were missing at the time of publication.
XX	
XX	Sequence 617 BP; 134 A; 198 C; 157 G; 128 T; 0 other;
SQ	
	Query Match            84.5%; Score 509.8; DB 22; Length 617;
	Best Local Similarity   92.0%; Pred. No. 1.2e-127;
	Matches   549; Conservative   0; Mismatches   47; Indels   1; Gaps   1;
OY	1 CAGGAACACCCTTCCTCGCCTGCTGCACCTGGACAATCAACTGCCAGCATGTC 60
Db	1 CAGGAACACCCTTCCTCGCCTGCTGCACCTGGACAATCAACTGCCAGCATGTC 60
OY	61 CTGCCAGCAAGAACAGCAGCAGTGTCCAAACCCCAACCCCAAGTGCTCCACCAAGTGTCC 120
Db	61 CTGCCAGCAAGAACAGCAGCAGTGTCCAAACCCCAACCCCAAGTGCTCCACCAAGTGTCC 120
OY	121 CCAGAAGACCCAGATACAGTGTCTGCTCCAGCTTCCTGAGTGTGGTGGCCCCAAGCTGTGG 180
Db	121 CCAGAAGACCCAGATACAGTGTCTGCTCCAGCTTCCTGAGTGTGGTGGCCCCAAGCTGTGG 180
OY	181 GGTCTGTGGCCTTAAGTCCAGAGGCGGTGTCTTTCTGAACACACACAGGCGCACCAATCCG 240
Db	181 GGTCTGTGGCCTTAAGTCCAGAGGCGGTGTCTTTCTGAACACACAGGCGCACCAATCCG 240
OY	241 ATGCGGGGGCAGAGGSCCAACTCTGTGACAGGGGAGTGTACAGAGGGGGGGCTC 300
Db	241 ATGCGGGGGCAGAGGSCCAACTCTGTGACAGGGGAGTGTACAGAGGGGGGGCTC 300
OY	241 ATGCGGGGGCAGAGGSCCAACTCTGTGACAGGGGAGTGTACAGAGGGGGGGCTC 300
Db	241 ATGCGGGGGCAGAGGSCCAACTCTGTGACAGGGGAGTGTACAGAGGGGGGGCTC 300
OY	301 TGCTGTCTGCCAGGTTCTGGGGGTGTGTGATCCAGTCTTGATGCTGAGCAAGCG 360
Db	301 TGCTGTCTGCCAGGTTCTGGGGGTGTGTGATCCAGTCTTGATGCTGAGCAAGCG 360
OY	301 TGCTGTCTGCCAGGTTCTGGGGGTGTGTGATCCAGTCTTGATGCTGAGCAAGCG 360
Db	301 TGCTGTCTGCCAGGTTCTGGGGGTGTGTGATCCAGTCTTGATGCTGAGCAAGCG 360
OY	361 ATCTTTGAGAGAAACAAGATGCCAAGAGGCAAGAACAGCCCATTAGACCATGCTT 420
Db	361 ATCTTTGAGAGAAACAAGATGCCAAGAGGCAAGAACAGCCCATTAGACCATGCTT 420
OY	361 ATCTTTGAGAGAAACAAGATGCCAAGAGGCAAGAACAGCCCATTAGACCATGCTT 420
Db	361 ATCTTTGAGAGAAACAAGATGCCAAGAGGCAAGAACAGCCCATTAGACCATGCTT 420
OY	421 CCCATAACCTCTTCTGACTTTCACAGGCTGAGCTGAGAGTTTTTCTGTGGGGATCTG 480
Db	421 CCCATAACCTCTTCTGACTTTCACAGGCTGAGCTGAGAGTTTTTCTGTGGGGATCTG 480
OY	420 CCCAGATACCTCTTCTGACTTTCACAGGCTGAGCTGAGAGTTTTTCTGTGGAGGTCTG 479
Db	420 CCCAGATACCTCTTCTGACTTTCACAGGCTGAGCTGAGAGTTTTTCTGTGGAGGTCTG 479
OY	481 AGCTCTCCCAAGAGCACTTCTGTTTATGTACAGATGTCAATATGTCTCCCTACCC 540
Db	481 AGCTCTCCCAAGAGCACTTCTGTTTATGTACAGATGTCAATATGTCTCCCTACCC 540
OY	480 AGCTCTCCCAAGAGCACTTCTGTTTATGTACAGATGTCAATATGTCTCCCTACCC 539
Db	480 AGCTCTCCCAAGAGCACTTCTGTTTATGTACAGATGTCAATATGTCTCCCTACCC 539
OY	541 TGTACTCCCAAGATGTGCACTGTTGTGCCCCAAGCTGTAAAAAAGATTAAGTTC 597
Db	541 TGTACTCCCAAGATGTGCACTGTTGTGCCCCAAGCTGTAAAAAAGATTAAGTTC 597







XX	RESULT 9
ID	AAS76924 standard; cDNA; 540 BP.
AC	AAS76924;
XX	
DT	13-FEB-2002 (first entry)
XX	
DE	DNA encoding novel human diagnostic protein #12728.
XX	
KM	Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS	Homo sapiens.
PX	MO200175067-A2.
PD	
XX	11-OCT-2001.
PF	30-MAR-2001; 2001WO-US068631.
PR	31-MAR-2000; 2000US-0540217.
XX	23-AUG-2000; 2000US-0649167.
PA	(HYSB-) HYSBO INC.
PI	Drmnac RT, Liu C, Tang YT,
DR	WPI; 2001-639362/73.
XX	P-PDSB; ABG12737.
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
XX	biodiversity -
PS	Claim 1, SEQ ID NO 12728; 103pp; English.
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (III) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AAs64197-AA84564 represent novel human
CC	diagnostic coding sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WFO
CC	at ftp.wipo.int/pub/published_pot_sequences.
XX	
XX	Sequence 540 BP; 105 A; 147 C; 143 G; 145 T; 0 other;
Query Match	15.3%; Score 95.6; DB 23; Length 540;
Best Local Similarity	69.3%; Pval. No. 6.7e-16;
Matches 147; Conservative	0; Mismatches 59; Indels 6; Gaps 1
DB	132 CAGTCAAGATGTCTGCCTTCAGCTTTCTTGCTGGCTGACCCAAAGCTCTGGAGTCTGTGACC 191
3	CAGGCCCATGCCACCTCCACGATCTCTTCGCTGTGTGTTTCAGACTCTGGGGGCTGTGCA 62
192	CTAAGCTCAGAGGGGGGCTGCTTCTTAACAACACAGGGCGCACCA-----CGATGCC 245

Db	63	GCCTGAGGATGATGCTGCTGCTGCTGAGCCACCAAGGCCCGCCGACCTGCTCGAAGCC	122
Qy	246	GGGGCCAGAGGGCCCAACTCTCTGTGACAGGGGCAAGTGTGACAGAGGGGGGCTTGAGCT	305
Db	123	GACCTCAGAGTTTCCAGCTGCTGTGGAGTGGCAATGGCCAGACTCTGGGGGCTTCAGCT	182
Qy	306	GCTGCCACGATTTCTGGGGGCTGCTGATCC	337
Db	183	GCTGCCACAGCTCTGGGGGCTGCTGCTGCTCC	214

RESULT 10  
AAK52164  
ID AAK52164 standard, cDNA, 414 BP.  
AAK52164  
AC  
AG  
AAK52164;  
06-NOV-2001 (first entry)  
XX  
XX  
DE Human polynucleotide SEQ ID NO 709.  
XX  
XX Human, cytokine; cell proliferation; cell differentiation; gene therapy;  
XX  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX  
XX nervous system disorder; arthritis; inflammation; ss.  
XX  
OS Homo sapiens.  
XX  
XX NC0200157190-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 05-FEB-2001; 2001WO-US04098.  
XX  
XX 03-FEB-2000; 2000US-0496914.  
XX  
XX 27-APR-2000; 2000US-0560875.  
XX  
XX 20-JUN-2000; 2000US-0598075.  
XX  
XX 19-JUL-2000; 2000US-0620325.  
XX  
XX 01-SEP-2000; 2000US-0654936.  
XX  
XX 15-SEP-2000; 2000US-0663561.  
XX  
XX 20-OCT-2000; 2000US-0693325.  
XX  
XX 30-NOV-2000; 2000US-0728422.  
XX  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Dymnac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
XX  
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
XX  
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
XX WPI: 2001-476283/51.  
XX  
XX P-PSDB; AAMW9031.  
XX  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
XX  
XX useful in diagnosis and gene therapy -  
XX  
XX  
XX Claim 1; Page 2436; 6221pp; English.  
XX  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
XX  
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
XX  
XX cytokine, cell proliferation or cell differentiation and which may induce  
XX  
XX production of other cytokines in other cell populations. The  
XX  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX  
XX activin/inhibin activity and may be useful in the diagnosis and/or  
XX  
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX  
XX inflammation.  
XX  
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
XX  
XX (AAM80020) are omitted as the relevant pages from the sequence listing  
XX  
XX were missing at the time of publication.  
XX





XX Homo sapiens.  
 OS  
 XX W0200061627-A1.  
 PN  
 XX 19-OCT-2000.  
 PD  
 XX 06-APR-2000; 2000WO-US09067.  
 PF  
 XX 09-APR-1999; 99US-0126697.  
 PR 20-JUN-2000; 2000US-0176829.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Rosen CA, Ruben SM, Komatsu G;  
 XX  
 XX WPI: 2000-647419/62.  
 DR P-PSDB; AAB45407.  
 XX  
 XX New nucleic acid molecules encoding 49 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives -  
 XX  
 XX Claim 1; Page 438; 521pp; English.

XX The invention relates to the isolation of genes A081710-C81758 encoding  
 CC 49 human secreted proteins AAB45386-B54534. The genes can be used to  
 CC generate fusion proteins by linking to the gene for the human  
 CC immunoglobulin G Fc portion (SEQID1) for increasing the stability of  
 CC the fusion protein as compared to the human protein only. The genes and  
 CC proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated  
 CC from a range of human tissues disclosed in the specification. The  
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)  
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 XX  
 XX Sequence 695 BP; 137 A; 230 C; 168 G; 157 T; 3 other;

Query Match 14.2%; Score 85.8; DB 21; Length 695;  
 Best Local Similarity 59.5%; Pred. No. 3.2e-13;  
 Matches 234; Conservative 1; Mismatches 98; Indels 60; Gaps 3;

QY 1 CAGGACAGAGCTTCTCTGCTCTCTGACCTGACCACTCACTCTGCGCAAGATGTC 60  
 DB 66 CAGGAGAGTGTGTGTGTCTCTGCTGTGACAGAGGTGACTAACTCTGCGCAAGATGTC 125  
 QY 61 CTGCGACCAAGACAG 107  
 DB 126 TTGCGACCAAGACAG 185  
 QY 108 -----CACCCAAAGTGTC 159  
 DB 186 AATATGTCACTAAGTGTC 245  
 QY 160 TG-----GCTGTGCGCCCAAGTCTTGGGT 183  
 DB 246 TGCAAGTCTTCT 305  
 QY 184 CTGTGAGCTTCT 243  
 DB 306 CTGTGAGCTTCT 365  
 QY 244 CCGGCGCAAG 303

DB 366 CCGGCGCC---GGACACAG 422  
 QY 304 CTGTGAGCTTCT 336  
 DB 423 CTGTGAGCTTCT 455

## RESULT 13

AAC85555  
 ID AAC85555 standard; cDNA; 526 BP.

XX AAC85555;  
 AC

DT 04-JUN-2001 (first entry)  
 XX

DE cDNA encoding CDIFP-14, Incyte ID No. 3432292GB1.

XX Human; cell differentiation; CDIFP; agonist; antagonist; epilepsy;  
 KW cell proliferation; Alzheimer's disease; schizophrenia disorder;  
 KW arteriosclerosis; cancer; atherosclerosis; diabetes mellitus; ss.  
 OS Homo sapiens.

XX Key

XX CDS Location/Qualifiers  
 FT 54..446  
 FT /tag= a  
 FT /product= "CDIFP-14"

XX W0200119860-A2.  
 XX

PD 22-MAR-2001.  
 XX

PP 14-SEP-2000; 2000WO-US25435.  
 XX

XX 15-SEP-1999; 99US-0154140.  
 PR

PR 06-DEC-1999; 99US-0169155.  
 XX

XX (INCY-) INCYTE GENOMICS INC.  
 PA

PI Tang YT, Hillman JL, Yue H, Reddy R, Lal P, Shah P, Azimzai Y;  
 PI Baughn WR, Lu DM, Bardman O, Shin JL, Patterson C;

DR WPI: 2001-211447/21.  
 DR P-PSDB; AAB47133.

XX Isolated polypeptides and polynucleotides involved in cell  
 PT differentiation are used for treatment, prevention and diagnosis of  
 PT cell proliferative, developmental and neurological disorders e.g.  
 PT cancer and Alzheimer's disease -

XX Claim 5; Page 127-128; 137pp; English.

XX The sequences given in AAC85545-72 encode human polypeptides involved  
 CC in cell differentiation (CDIFP). CDIFP polypeptides and agonists of  
 CC these are used to treat a disease or condition associated with  
 CC decreased expression of functional CDIFP. An antagonist of CDIFP is  
 CC used to treat a disease or condition associated with over expression  
 CC of functional CDIFP. CDIFP polypeptides may be used for the treatment,  
 CC prevention and diagnosis of cell proliferative, developmental and  
 CC neurological disorders, such as Alzheimer's disease, schizophrenia  
 CC disorders, arteriosclerosis, cancer, atherosclerosis, diabetes mellitus  
 CC and epilepsy. The CDIFP-14 sequence is homologous to human skin  
 CC specific protein.  
 CC

XX Sequence 526 BP; 96 A; 176 C; 139 G; 115 T; 0 other;

QY Query Match 12.0%; Score 72.4; DB 22; Length 526;  
 Best Local Similarity 59.0%; Pred. No. 1.2e-09;  
 Matches 209; Conservative 0; Mismatches 96; Indels 49; Gaps 3;

QY 30 ACCTGACAACTCACT 89





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 05:39:35 ; Search time 2280 Seconds

(without alignments)  
6427.894 Million cell updates/sec

Title: US-09-744-315-7

Perfect score: 603

Sequence: 1 ccgagacagccttcctcctgc.....aaagataagcttcctgc 603

Scoring table: IDENTITY NUC

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:  
1: em\_estba:\*  
2: em\_estbm:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_iny:\*  
19: em\_gss\_pih:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_pbg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	468	77.6	477	13	BQ300994 PMO-KT004
C 2	409.8	68.0	528	9	AM136771 UI-R-B11-
C 3	398.8	65.1	514	9	AM205844 UI-R-B11-
C 4	321.4	53.3	470	12	B1670520 SPRL 6b H

5	317.4	52.6	470	12	B1670519	B1670519 SPRL 6a H
C 6	286	47.4	407	9	AM137411	AM137411 UI-R-B11-
C 7	189.6	31.4	567	12	BM389983	BM389983 UI-R-CN1-
C 8	186	30.8	485	14	BY708906	BY708906 BY708906
C 9	185	30.6	570	11	AK009036	AK009036 Mus muscu
C 10	184.4	30.7	231	10	BP354854	BP354854 MR2-HT078
C 11	184.4	30.6	757	14	CB575489	CB575489 AGENCOURT
C 12	182.4	30.2	563	11	AK008201	AK008201 Mus muscu
C 13	182.4	30.2	564	11	AK009001	AK009001 Mus muscu
C 14	182.4	30.2	944	14	CB590520	CB590520 AGENCOURT
C 15	182.2	30.2	543	12	BI277681	BI277681 UI-R-C20-
C 16	175.6	29.1	429	14	W34356	W34356 ma9a09.x1
C 17	174.6	29.0	535	14	CB575977	CB575977 AGENCOURT
C 18	172	28.5	627	14	CB589787	CB589787 AGENCOURT
C 19	171.2	28.4	401	9	AA059561	AA059561 t65f03.r
C 20	171.2	28.4	537	4	BM523362	BM523362 RZPD Mus
C 21	171.2	28.4	638	14	CB588625	CB588625 AGENCOURT
C 22	169.2	28.1	624	14	CB321483	CB321483 AGENCOURT
C 23	166.4	27.6	606	14	CB575565	CB575565 AGENCOURT
C 24	165.4	27.4	552	12	BI285481	BI285481 UI-R-DD0-
C 25	161.2	26.7	601	9	AV071932	AV071932 AV071932
C 26	153.4	25.4	347	14	W18301	W18301 mb87f09.r1
C 27	148.6	24.6	203	12	BI670515	BI670515 SPRL 3a H
C 28	142.6	23.6	307	10	BF545883	BF545883 UI-R-C2-m
C 29	142	23.5	227	10	BE717933	BE717933 MR2-HT078
C 30	138.6	23.0	487	12	BI286064	BI286064 UI-R-DD0-
C 31	133.8	22.2	477	10	BF417300	BF417300 UI-R-CNO-
C 32	119.4	19.8	401	14	W37034	W37034 mb72b10.x1
C 33	111.8	18.5	432	12	BM384766	BM384766 UI-R-CN1-
C 34	111.4	18.5	394	12	BI670517	BI670517 SPRL 4a H
C 35	109.8	17.2	252	14	W11697	W11697 mb21a07.r1
C 36	103.6	16.6	571	12	BI670518	BI670518 SPRL 5a H
C 37	98.6	16.4	282	10	BB565509	BB565509 BB565509
C 38	96.8	16.1	460	11	AK009081	AK009081 Mus muscu
C 39	96.6	16.0	220	8	BY355882	BY355882 BY355882
C 40	96.6	16.0	220	28	BH041328	BH041328 RPCI-24-3
C 41	95.6	15.9	540	9	AW291010	AW291010 UI-R-B12-
C 42	95.6	15.9	832	10	BG198345	BG198345 RST17603
C 43	92	15.3	271	10	BB565495	BB565495 BB565495
C 44	90.4	15.0	258	10	BB565531	BB565531 BB565531
C 45	89	14.8	248	10	BB565379	BB565379 BB565379

## ALIGNMENTS

### RESULT 1

BQ300994/c 477 bp mRNA linear EST 16-MAY-2002  
LOCUS FMO-KT0041-030201-003-b12 KT0041 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BQ300994  
ACCESSION BQ300994  
VERSION BQ300994.1 GI:20816516  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

### REFERENCE

1 (bases 1 to 477)  
Mammalia: Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsumura, A., Bata, G.S., Simpson, D.H.,  
Brunstein, A., deoliveira, P.S., Bucher, P., Jorgensen, C.V., O'Hare  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
JOURNAL MEDLINE  
PUBMED 10737800  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL:  
(http://www.ludwig.org.br/scripts/gethtml2.pl?i=PM0&t2=PM0-KT0041-030201-003-b1&t&3=2001-02-03&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 476.

## FEATURES

source

Location/Qualifiers

1..477

/organism="Homo sapiens"

/mol\_type="RNA"

/db\_xref="taxon:9606"

/dev\_stage="Adult"

/clone\_lib="KT0041"

/note="Organ: bladder tumor; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 101 a 127 c 156 g 92 t 1 others

ORIGIN

Query Match

Best Local Similarity 98.7%; Pred. No. 5,4e-113;

Matches 471; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

89 CCCCACCCAGAGTCCCTCAGCCAGTGTCCCAAGAGCCCAAGTACAGTGTCCCT 148

477 CCCCACCCAGTGTCCCTCAGCCAGTGTCCCAAGAGCCCAAGTACAGTGTCCCT 418

149 CCAAGTCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 208

417 CCAAGTCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358

209 TGTCTTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 268

357 TGTCTTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 298

269 GACAGGAGGAGT 328

297 GACAGGAGGAGT 238

329 TGTCTTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 388

237 TGTCTTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 178

389 GACAGGAGGAGT 448

177 GACAGGAGGAGT 118

449 GGTGAGCTGAGAGT 508

117 GGTGAGCTGAGAGT 58

509 TATGTACAGAGT 565

57 TATGTACAGAGT 1

Db

RESULT 2

AM136771/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

Homo sapiens (human)

528 bp mRNA linear EST 29-OCT-1999

UI-H-B1-adm-f-02-0-UI-s1 NCI\_CGAP\_Sub3 Homo sapiens cDNA clone

IMAGE:2717330 3', mRNA sequence.

AM136771

AM136771.1 GI:6140904

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 528)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

The sequence contained an oligo-dt track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:

NCI-CGAP clone distribution information can be found through the

I.M.A.G.E. Consortium/ILNL at:

www-bio.lnl.gov/bdbp/image/image.html

Seq primer: M13 forward

POLYA=yes.

FEATURES

source

Location/Qualifiers

1..528

/organism="Homo sapiens"

/mol\_type="RNA"

/db\_xref="taxon:9606"

/clone\_lib="IMAGE:2717330"

/lab\_name="DH10B (Life Technologies)"

/note="Vector: pTZ19-Pac (pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; The

NCI CGAP Sub3 library is a subtracted library derived from

the NCI CGAP Sub1 library, which is a subtracted library

derived from B1. B1 constitutes a mixture of 21

normalized or subtracted NCI CGAP libraries: NCI\_CGAP\_Co4

, NCI\_CGAP\_Pr22, NCI\_CGAP\_Pr28, NCI\_CGAP\_Co10,

NCI\_CGAP\_Co16, NCI\_CGAP\_Kid5, NCI\_CGAP\_Kid12,

NCI\_CGAP\_Kid3, NCI\_CGAP\_Co8, NCI\_CGAP\_Cli1, NCI\_CGAP\_Le12,

NCI\_CGAP\_Br2, NCI\_CGAP\_Co8, NCI\_CGAP\_Cli1, NCI\_CGAP\_Le12,

NCI\_CGAP\_Br23, NCI\_CGAP\_Lu5, NCI\_CGAP\_Lu24,

NCI\_CGAP\_Lu19, NCI\_CGAP\_Gc4, NCI\_CGAP\_Gc6,

NCI\_CGAP\_Br25. These 21 libraries were pooled and a

single-stranded DNA preparation of the resulting mixture

was used as a tracer in a subtractive hybridization with

a driver whose composition is detailed below:

NCI\_CGAP\_Kid3 pool 1 LLM 3334-3337, 3682-3683,

3798-3803 (IMAGE Cloneids 122376-132391), 1456008-1456775

, 1500552-1502855; NCI\_CGAP\_Kid5 pool 1 LLM 3338-3342

, 3722-3725, 3776-3778 (IMAGE Cloneids 123912-132831,

1471368-1472903, 1492104-1493255); NCI\_CGAP\_Lu5 pool 1

LLM 3575-3582, 3851-3854 (IMAGE Cloneids 144970-1417991,

1520904-1522439); NCI\_CGAP\_Gc4 pool 1 LLM 3164-3167,

3716-3720, 3733-3735 (IMAGE Cloneids 1257096-1258631,

1469064-1470983, 1475592-1476743); NCI\_CGAP\_Pr22 pool 1

LLM 2457-2459, 2758-2759, 3062-3068 (IMAGE Cloneids

985608-986759, 1101193-1101959, 1217928-1220615);

NCI\_CGAP\_Co10 pool 1 LLM 2644-2653, 2871-2872 (IMAGE

Cloneids 1057416-1061255, 1145844-1145351). Subtraction

was performed as previously described [Bonardo-Lemon &amp;

Soares (1996): Normalization and Subtraction: Two

Approaches to Facilitate Gene Discovery. Genome Research

6, 791-806.

TAG LIB=NCI\_CGAP\_Br26

TAG TISSUE=Brain

TAG\_SEQ=ATAG3"

BASE COUNT 114 a 157 g 115 t 2 others

ORIGIN

Query Match

Best Local Similarity 90.2%; Pred. No. 1.4e-97;

Matches 449; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

100 GTGTCCTCAGCAGAGTGTCCCAAGAGCCCAAGTACAGTGTCTCCAGTCTCTC 159

477 GTGTCCTCAGCAGAGTGTCCCAAGAGCCCAAGTACAGTGTCTCTCCAGTCTCTC 469

Db



Db 35 AACCTCC 29

RESULT 4  
LOCUS B1670520 470 bp mRNA linear EST 12-SEP-2001  
DEFINITION SPRL 6b Human SPRL genes Homo sapiens cDNA, mRNA sequence.  
ACCESSION B1670520  
VERSION B1670520.1 GI:15585904  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Wang, A.  
TITLE Human SPRL genes  
JOURNAL Unpublished  
COMMENT Contact: Aijin Wang  
Department of Carcinogenesis  
MD Anderson Cancer Center, University of Texas  
Smithville, TX 78957, USA  
Email: awang@sprd1.mdacc.tmc.edu.

FEATURES  
source  
1..470  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone\_lib="Human SPRL genes"  
/note="Cloned by RT-PCR"

BASE COUNT 93 a 157 c 125 g 95 t

ORIGIN

Query Match 53.3%; Score 321.4; DB 12; Length 470;  
Best Local Similarity 94.1%; Pred. No. 3.2e-74;  
Matches 334; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 19 GCCTCTCTGCACTGAGCAACTCACTCTGCAAGATGTCCTGACAGACAGACAGA 78  
Db 115 GCTTTGTTTGTTCAGAGACAATCACTCTGCAAGATGTCCTGACAGACAGACAGA 174  
QY 79 GCAATGCCAACCCCAACCAAGTGTCTCTGACCAAGTGTCTCTGACCAAGTGTCT 138  
Db 175 GCAATGCCAACCCCAACCAAGTGTCTCTGACCAAGTGTCTCTGACCAAGTGTCT 234  
QY 139 GTGCTGCTGCACTGAGCAACTCACTCTGCAAGATGTCCTGACAGACAGACAGA 198  
Db 235 GTGCTGCTGCACTGAGCAACTCACTCTGCAAGATGTCCTGACAGACAGACAGA 294  
QY 199 CGAGGGGCGGCTGCTCTCTGAAACCAACAGGGGCGCAACCGATGCGGGCGCCAGAGGCC 258  
Db 295 CGAGGGGCGGCTGCTCTCTGAAACCAACAGGGGCGCAACCGATGCGGGCGCCAGAGGCC 354  
QY 259 CAATCTCTGTGACAGGGGCGAGTGTGACAGAGGGGCGCTCTGGCTGTGCGCCAGGCTC 318  
Db 355 CAATCTCTGTGACAGGGGCGAGTGTGACAGAGGGGCGCTCTGGCTGTGCGCCAGGCTC 414  
QY 319 TGGGAGGCTGCTGCTGATCCAGATCTGATGCTGAGCAAGCGATCTTTTGAGAGAA 373  
Db 415 TGGGAGGCTGCTGCTGATCCAGATCTGATGCTGAGCAAGCGATCTTTTGAGAGAA 469

RESULT 5  
LOCUS B1670519 470 bp mRNA linear EST 12-SEP-2001  
DEFINITION SPRL 6a Human SPRL genes Homo sapiens cDNA, mRNA sequence.  
ACCESSION B1670519  
VERSION B1670519.1 GI:15585903  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 470)  
AUTHORS Wang, A.  
TITLE Human SPRL genes  
JOURNAL Unpublished  
COMMENT Contact: Aijin Wang  
Department of Carcinogenesis  
MD Anderson Cancer Center, University of Texas  
Smithville, TX 78957, USA  
Email: awang@sprd1.mdacc.tmc.edu.

FEATURES  
source  
1..470  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone\_lib="Human SPRL genes"  
/note="Cloned by RT-PCR"

BASE COUNT 93 a 156 c 125 g 96 t

ORIGIN

Query Match 52.6%; Score 317.4; DB 12; Length 470;  
Best Local Similarity 95.3%; Pred. No. 3.6e-73;  
Matches 327; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 31 CCTGCAACTCACTGAGCAACTCACTGAGCAAGTGTCTCTGACAGACAGACAGA 90  
Db 127 CAGGACCACTCACTGAGCAACTCACTGAGCAAGTGTCTCTGACAGACAGACAGA 186  
QY 91 CCAACCAAGTGTCTCTGACCAAGTGTCTCTGACCAAGTGTCTCTGACCAAGTGTCT 150  
Db 187 CCAACCAAGTGTCTCTGACCAAGTGTCTCTGACCAAGTGTCTCTGACCAAGTGTCT 246  
QY 151 AGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 210  
Db 247 AGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 306  
QY 211 CTCTCTGAAACCAACAGGGGCGCAACCGATGCGGGGCGCAAGGCCCACTCTGTA 270  
Db 307 CTCTCTGAAACCAACAGGGGCGCAACCGATGCGGGGCGCAAGGCCCACTCTGTA 366  
QY 271 CAGGGGCACTGATGAGCAAGGGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 330  
Db 367 CAGGGGCACTGATGAGCAAGGGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 426  
QY 331 CTGATCCAGATCTGATGCTGAGCAAGCGATCTTTTGAGAGAA 373  
Db 427 CTGATCCAGATCTGATGCTGAGCAAGCGATCTTTTGAGAGAA 469

RESULT 6  
LOCUS AM137411/c 407 bp mRNA linear EST 29-OCT-1999  
DEFINITION UI-H-B11-acs-e-01-0-UI-s1 NCI CGAP\_Skb3 Homo sapiens cDNA clone  
IMAGE:2715360 3', mRNA sequence.  
ACCESSION AM137411  
VERSION AM137411.1 GI:61411729  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:  
NCI-CGAP clone distribution information can be found through the

I.M.A.G.E. Consortium/ILNL at:  
www.bio.lnln.gov/bbrp/image/image.html  
Seq primer: M13 forward  
POLYA=yes

## FEATURES

source

Location/Qualifiers

1..407

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:2715360"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pTR73D-Pac (Pharmacia) with a modified

polylinker. Site 1: Not I; Site 2: Eco RI; The

NCI CGAP Subj library is a subtracted library derived from

the NCI CGAP Subj library, which is a subtracted library

derived from B1. B1 constitutes a mixture of 21

normalized or subtracted NCI CGAP libraries: NCI CGAP Co4

, NCI CGAP Pr22, NCI CGAP Pr28, NCI CGAP Co10,

NCI CGAP Co16, NCI CGAP Kid5, NCI CGAP Kid12,

NCI CGAP Kid3, NCI CGAP Kid11, NCI CGAP Lym2,

NCI CGAP Br2, NCI CGAP Co8, NCI CGAP C11, NCI CGAP Le12,

NCI CGAP Brn23, NCI CGAP Lys, NCI CGAP Lu24,

NCI CGAP Lu19, NCI CGAP Gc4, NCI CGAP Gc6,

NCI CGAP Brn25. These 21 libraries were pooled and a

single-stranded DNA preparation of the resulting mixture

was used as a tracer in a subtractive hybridization with

a driver whose composition is detailed below:

NCI CGAP Kid3 pool 1 LLM 3334-3337, 3682-3683,

3798-3803 (IMAGE Clonoids 132376-132391, 145608-145675

, 150052-150285); NCI CGAP Kid5 pool 1, LLM 3338-3342

, 3722-3725, 3776-3778 (IMAGE Clonoids 132391-132581,

1471368-1472903, 1492104-1493255); NCI CGAP Lys pool 1,

LLM 3575-3582, 3851-3854 (IMAGE Clonoids 1414920-1417991,

1520904-1522439); NCI CGAP Gc4 pool 1 LLM 3164-3167,

3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631,

1469064-1470983, 1475592-1476743); NCI CGAP Pr22 pool 1

LLM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids

985608-986759, 1101192-1101959, 1217928-1220615);

NCI CGAP Co10 pool 1 LLM 2644-2653, 2871-2872 (IMAGE

Clonoids 1057416-1061255, 1144584-1145351). Subtraction

was performed as previously described [Bonaldi, Lennon &amp;

Soares (1996): Normalization and Subtraction: Two

Approaches To Facilitate Gene Discovery. Genome Research

6: 791-806.

TAG LIB=NCI CGAP Lu19

TAG\_TISSUE=Lung

TAG\_SEQ=GACGAC

BASE COUNT 87 a 107 c 113 g 100 t

ORIGIN

Query Match 47.4%; Score 286; DB 9; Length 407;

Best Local Similarity 87.6%; Pred. No. 7e-65;

Matches 324; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

228 GGGGCGACCAAGGATCCGGCGCCAGAGGCCCAACTCTCTGACAGGGGCGATGCTTACG 287

239 GGGGCGACCAAGGATCCGGCGCCAGAGGCCCAACTCTCTGACAGGGGCGATGCTTACG 338

288 AAGCGGGGGCTCTGGCTCTGCGACGGTTCTGGGGGGCTGCTGATGATCCAGATCTGAT 347

337 AAGCGGGGGCTCTGGCTCTGCGACGGTTCTGGGGGGCTGCTGATGATCCAGATCTGAT 278

348 GCTGAACCAAGCATTTTGAAGAAACAAGATCCCAAGAGGCCCAAGACAGCCCATC 407

277 GCTGAACCAAGCATTTTGAAGAAACAAGATCCCAAGAGGCCCAAGACAGCCCATC 219

408 TGAAGCATGCTTCCATATACCTCTTGAATTTCACAGGCTAGCTGAGGTTTTC 467

218 TGAAGCATGCTTCCATATACCTCTTGAATTTCACAGGCTAGCTGAGGTTTTC 159

468 TGTGGGGGATCTAGCTTCTCCCAAGAGCATCTTGTGTTTATGATGATGATCAT 527

Db 158 TGTGAAGCTGTAGCTCTCCCAAGAGCATCTTCTGTATGTTCAGATGTACAT 99

QY 528 GTCCCTTACCCCTGTACTGTCCAGAGATTTGGAGTGTTCGCCCACTCTTAATAA 587

Db 98 GTCCCTTACCCCTGTACTGTCCAGAGATTTGGAGTGTTCGCCCACTCTTAATAA 39

QY 588 GATTAAGTTC 597

Db 38 ATTAAGTTC 29

RESULT 7 567 bp mRNA linear EST 17-JUN-2002

BM389983

UT-R-CNI-cjo-g-12-0-UI s1 UT-R-CNI Rattus norvegicus cDNA clone

LOCUS UT-R-CNI-cjo-g-12-0-UI 3', mRNA sequence.

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murinae;

Rattus.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBR, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence contained an oligo-dr track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonaldi poly A

tail. cDNA library preparation: M.B. Soares Lab Clone distribution:

clones will be available through Research Genetics (www.resgen.com)

Seq primer: M13 forward

POLYA=yes

Location/Qualifiers

1..567

/organism="Rattus norvegicus"

/mol\_type="mRNA"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-CNI-cjo-g-12-0-UI"

/dev\_stage="adult"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="UI-R-CNI"

/note="Vector: pTR73D-Pac (Pharmacia) with a modified

polylinker. Site 1: Not I; Site 2: Eco RI; The UI-R-CNI

library is a subtracted library derived from the following

pool of seven normalized rat libraries: normalized rat

seminal vesicles, normalized rat penis, normalized rat

bladder, normalized rat cervix, and normalized rat

adipose, normalized rat fundus, and normalized rat

salivary gland. It was constructed according to the

procedure described by Bonaldi, Lennon &amp; Soares (Genome

Research 6: 791-806, 1996). For construction of

the CNI library, plasmid DNA from the pool of normalized

libraries was electroporated into competent bacteria for

the production of single-stranded circular DNA. This was

then used as a tracer in a subtractive hybridization with

a driver (PCR amplified inserts from a plasmid DNA template

preparation) comprising: a) a pool of about 34,000 clones

from the Rat Unigene Set corresponding to plates R-5-AA-NN

excluding plates R-5-MM and MM. This pool represented 40%









DNA Sequencing by: Agencourt Bioscience Corporation

Mus musculus  
Eukaryota: Metazoa: Chordata: Cr

SOURCE Mus musculus (house mouse)  
ORDINARY Mus musculus

ORGANISM      Mus musculus  
Eukaryota: Metazoa: Chordata: Mammalia: Rodentia: Muridae: Murinae: Murini: Mus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



Db 265 CACAGATGACGGGCGAGAGCTCCAGTTCTGTGACCGTGGCAGGGTCAGAGCTGGG 324

Qy 296 GGCTCGGCGCTGCGCCAGGTTCTGGGGGCTGCTGATCAGATCCTGATGCTGAGAC 355

Db 325 GGGTCAGGCGTGTGGCCACAGCTCTGGGAGCTGCTCTACCTGGAGCATATGTGGAGAC 384

Qy 356 AAGCAGCTTTGGAGGAAACAGATCCCAAGAGCCAAAGACCCC 403

Db 385 AAGACACGTCTGGAGGAATCAACA-TCCACAGGGGTCAAGAAAGCCC 429

Accession	Definition	LOCUS	AK009001	RESULT 13
AK009001	Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone2300002C02 product:ERG3 (ERG3 PROTEIN) homolog [Mus musculus], full insert sequence.	564 bp	mRNA	linear
				HTC 05-DEC-2002

REFERENCE  
1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)

**TITLE** Functional annotation of a full-length mouse cDNA collection  
**JOURNAL** Nature 409 (6821), 685-690 (2001)

```

DAM330 row: c column: 02
lity sequence stop: 550.
location/Qualifiers
1..944
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30290065"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_id="N1H MCC 136"
/notes="Vector: pCMV-SPORT6.1.ccdB, Site 1: EcoRV, Site 2
NotI; Normalized, full-length enriched library from pool
of mouse embryonic limb, maxilla and mandible, embryonic

```

Email: bento-soares@iowa.edu  
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the non-normalized penis library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLY(A)yes.

# FEATURES

Source Location/Qualifiers  
1..543  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-C20-bp-f-12-0-UI"  
/dev\_stage="ADULT"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-C20"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C20 library is a non-normalized library constructed from rat penis tissue. For a detailed description of the library from which this clone was derived, please visit our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)  
TAG LIB=UI-R-C20  
TAG TISSUE=penis  
TAG\_SEQ=TTGCGAGACA"  
BASE COUNT 118 a 144 c 162 g 119 t  
ORIGIN

Query Match 30.2%; Score 182.2; DB 12; Length 543;  
Best Local Similarity 76.4%; Pred. No. 2.2e-37;

Matches 282; Conservative 0; Mismatches 68; Indels 19; Gaps 4;

OY	46	TTCTGCCAAGATGCTGCGAGCAGCAACGACAGTGCACACCCCAAGGATGCC	105
DB	527	TCATGCCAAGATGCTGCGAGCAGCAACGACAGTGCACACCTCTCCAAATGCC	468
OY	106	CTCAACC-----AAGTGTCCCAAGAGCCCAATGCTGCTCCAGC	153
DB	467	CTCCCCAAGTGTCTCCCAAGTGTCCCAAGAGCAGCAATGTCTGCTGCAGC	408
OY	154	TTCTGTGCTGTGCGCCCAAGCTCTGAGGCTCT--GTGGCCCTAGCTCCAGAGGCGCTG	210
DB	407	TTCTTCTGCTGTGCTCAAGCTCTGAGGCTCTGAGCTTCCAGCTCTGAGGAGGCTG	348
OY	211	CTTCTGAGCAGCAGCGGC--CAACACGATGCCGCGCAGAGGCCCACTCTG	267
DB	347	CTGCTGAGCAGCAGCGCGAGGCTCCCAAGATGACGCGCAGAGCTCCAGTTCTG	288
OY	268	TGACAGGAGGAGTGTGAGCAGAGCGGAGGCTGTGCTGCTGCAAGTTCTGAGGAGCTG	327
DB	287	TGACCTGTGAGTGTGAGCAGTCTGAGGAGCTGAGCTGTGCTGCAAGTTCTGAGGAGCTG	228
OY	328	CTGCTGATCCAGATCTGATGCTGAGCAAGCATCTTTGAGAGAAACAAATCCCAAG	387
DB	227	CTGCTAACCTGAGAGCATGATGCTGAGCATGAGA-GTCTGAGAGAAAGCATATCCCAAG	169
OY	388	AGGCCAAGA 396	
DB	168	GCTCAAGA 160	

Search completed: February 16, 2004, 07:54:52  
Job time : 2284 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 02:57:33, Search time 3257 Seconds

(Without alignments)  
7573.996 Million cell updates/sec

Title: US-09-744-315-7

Perfect score: 603

Sequence: 1 caggaacagccttcctcgc.....aaagaataagtcctcgc 603

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database:
1:  gb_ba:*
2:  gb_ba:*
3:  gb_ba:*
4:  gb_ba:*
5:  gb_ba:*
6:  gb_ba:*
7:  gb_ba:*
8:  gb_ba:*
9:  gb_ba:*
10: gb_ba:*
11: gb_ba:*
12: gb_ba:*
13: gb_ba:*
14: gb_ba:*
15: gb_ba:*
16: gb_ba:*
17: gb_ba:*
18: gb_ba:*
19: gb_ba:*
20: gb_ba:*
21: gb_ba:*
22: gb_ba:*
23: gb_ba:*
24: gb_ba:*
25: gb_ba:*
26: gb_ba:*
27: gb_ba:*
28: gb_ba:*
29: gb_ba:*
30: gb_ba:*
31: gb_ba:*
32: gb_ba:*
33: gb_ba:*
34: gb_ba:*
35: gb_ba:*
36: gb_ba:*
37: gb_ba:*
38: gb_ba:*
39: gb_ba:*
40: gb_ba:*
41: gb_ba:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	561.2	93.1	52355	AL1356426	AL1356426 Human DNA
2	511.4	84.8	617	BD181469	BD181469 A gene of
3	511.4	84.8	617	AB048288	AB048288 Homo sapi
4	480.4	79.7	76369	AC023208	AC023208 Homo sapi
5	465.4	77.2	76369	AC023208	AC023208 Homo sapi
6	360.4	59.8	52355	AL1356426	AL1356426 Human DNA
7	360.4	59.8	52355	AL1356426	AL1356426 Human DNA
8	198.6	31.1	199220	AC140641	AC140641 Macaca mu
9	187.6	30.9	275	BD181458	BD181458 A gene of
10	186.2	30.7	195503	AC137251	AC137251 Rattus no
11	185.4	30.7	261316	AC095371	AC095371 Rattus no
12	185.4	30.7	261316	AC095371	AC095371 Rattus no
13	185.4	30.7	347801	AC118786	AC118786 Rattus no
14	184.4	30.6	566	AF176515	AF176515 Mus muscu
15	184.4	30.6	566	AF176515	AF176515 Mus muscu
16	183.8	30.5	195503	AC137251	AC137251 Rattus no
17	183.8	30.5	311774	AC118786	AC118786 Rattus no
18	183.8	30.5	347801	AC118786	AC118786 Rattus no
19	181	30.0	3139	AF303430	AF303430 Mus muscu
20	177.2	29.4	199220	AC135289	AC135289 Mus muscu
21	169.4	28.1	213540	AC118873	AC118873 Rattus no
22	167.2	27.7	311774	AC118786	AC118786 Rattus no
23	142.6	23.6	689	AX525830	AX525830 Sequence
24	142.6	23.6	689	AX525830	AX525830 Sequence
25	134.4	22.3	261316	AC095371	AC095371 Rattus no
26	103.6	17.2	188285	AL135842	AL135842 Human DNA
27	82.8	13.7	96013	AL135379	AL135379 Human DNA
28	72.6	12.0	619	HUM2D48C09	AF086289 Homo sapi
29	72.6	12.0	619	HUM2D48C09	AF086289 Homo sapi
30	72.6	12.0	619	HUM2D48C09	AF086289 Homo sapi
31	68.8	11.4	116067	AL162596	AL162596 Homo sapi
32	68.8	11.4	265938	AC127247	AC127247 Mus muscu
33	67.2	11.1	231037	AC114064	AC114064 Rattus no
34	67.2	11.1	231037	AC114064	AC114064 Rattus no
35	67.2	11.1	293137	AC110356	AC110356 Rattus no
36	66.8	11.1	126610	AC132604	AC132604 Mus muscu
37	66.8	11.1	126610	AC132604	AC132604 Mus muscu
38	62.6	10.4	1210	BC038391	BC038391 Homo sapi
39	62.6	10.3	96013	AL135379	AL135379 Human DNA
40	62.6	10.3	258785	AC102892	AC102892 Rattus no
41	59.8	9.9	1197	BC031811	BC031811 Homo sapi
42	59.2	9.8	188285	AL135842	AL135842 Human DNA
43	58	9.6	557	G55179	G55179 SHGC-100235
44	52.2	8.7	544	AY113557	AY113557 Drosophila
45	51.8	8.6	265938	AC127247	AC127247 Mus muscu

## ALIGNMENTS

RESULT 1  
AL1356426/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AL1356426 52355 bp DNA linear PRI 15-NOV-2001  
Human DNA sequence from clone Rpl-11013 on chromosome 1q21.1-21.3,  
complete sequence.  
AL1356426 GI:16972880  
HTG.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 52355)  
Johnson, C.  
Direct Submission

## JOURNAL

Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
[hunquerry@sanger.ac.uk](mailto:hunquerry@sanger.ac.uk) Clone request: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
 On Nov 16, 2001 this sequence version replaced gi:116030120.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>  
 RPI-11013 is from the library RPI-1 constructed by the group of P. de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RPI-11013. It may be shorter because we sequence overlapping sections only once, except for a short overlap.  
 The true left end of clone RPI-5210 is at 50356 in this sequence. The true right end of clone RPI-915 is at 2000 in this sequence.  
 Location/Qualifiers

1..52355

/organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

/chromosome="1"

/map="q21.1-21.3"

/clone="RPI-11013"

/clone\_id="RPI-1"

BASE COUNT 15094 a 10374 c 10652 g 16235 t

## ORIGIN

Query Match 93.1%; Score 561.2; DB 9; Length 52355;  
 Best Local Similarity 97.8%; Pred. No. 2,7e-124;  
 Matches 569; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

19 GCTCTCTGCACTGGAACAATCACTCTGCGCAAGATGCTGCGCAAGAACGACGA 78  
 16087 GCTTGTCTTCTCAAGCAATCACTCTGCGCAAGATGCTGCGCAAGAACGACGA 16028  
 79 GCAGTGCACACCCCAACCAAGTCTGCGCAAGATGCTGCGCAAGAACGACGA 138  
 16027 GCAGTGCACACCCCAACCAAGTCTGCGCAAGATGCTGCGCAAGAACGACGA 15968  
 139 GTGTCTGCTCAAGTCTCTGCTGCTGCGCAAGTCTGCGCAAGTCTGCGCAAGTCT 198  
 15967 GTGTCTGCTCAAGTCTCTGCTGCTGCGCAAGTCTGCGCAAGTCTGCGCAAGTCT 15908  
 199 CGAGGCGCGTCTCTGCGCAAGTCTGCGCAAGTCTGCGCAAGTCTGCGCAAGTCT 258  
 15907 CGAGGCGCGTCTCTGCGCAAGTCTGCGCAAGTCTGCGCAAGTCTGCGCAAGTCT 15848  
 259 CAATCTCTGCAAGGCGCAAGTCTGCGCAAGTCTGCGCAAGTCTGCGCAAGTCT 318  
 15847 CAATCTCTGCAAGGCGCAAGTCTGCGCAAGTCTGCGCAAGTCTGCGCAAGTCT 15788  
 319 TGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 378

Db 15787 TGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 15728

QY 379 AATCCCAAGAGGCGCAAGAACCAAGCCCATCTGACGATGCTTCCATATACCTCTTG 438

Db 15727 AATCCCAAGAGGCGCAAGAACCAAGCCCATCTGACGATGCTTCCATATACCTCTTG 15668

QY 439 ACTTTACAGGCTGACCTGAGAGTTTCTGCGGAGATCTGAGCTTCCCAAGAGCA 498

Db 15667 ACTTTACAGGCTGACCTGAGAGTTTCTGCGGAGATCTGAGCTTCCCAAGAGCA 15608

QY 499 CTCTCTGTTTATGACAGATGTCATATGTCCTTACCTGATACCTGCAAGAGATG 558

Db 15607 CTCTCTGTTTATGACAGATGTCATATGTCCTTACCTGATACCTGCAAGAGATG 15548

QY 559 GCAGTCTTGTGCGCAAGTCTGTAAGATTAAGTTCTG 600

Db 15547 GCAGTCTTGTGCGCAAGTCTGTAAGATTAAGTTCTG 15506

## RESULT 2

BD181469

LOCUS BD181469 617 bp DNA linear PAT 15-MAY-2003

DEFINITION A gene of which expression changes in psoriasis and a method for e

AMBIENT BD181469

ACCESSION BD181469.1 GI:30792387

VERSION JP 2002330770-A/17.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 617)  
 Imai, Y., Wakimoto, K., Yamada, E., Chiba, H. and Okubo, K.  
 A gene of which expression changes in psoriasis and a method for e

JOURNAL xamination directed to said gene  
 Patent: JP 2002330770-A 17 19-NOV-2002;  
 TANABE SEIYAKU CO LTD

COMMENT OS Homo sapiens (human)  
 PN JP 2002330770-A/17  
 PD 19-NOV-2002

PF 25-MAY-2001 JP 2001156529

PI YUT IMAI, KOJI WAKIMOTO, ERIKO YAMADA, HIROAKI CHIBA, KOSAKU PI

OKUBO

PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21 PC

, C12N5/10, C12Q1/68,  
 PC G01N33/53, G01N33/53, G01N33/56, C12N15/00, C12N5/00 CC A gene

of which expression changes in psoriasis and a method CC

CC xamination directed to said gene

FT Key Location/Qualifiers  
 (58) . . (333).

FT CDS Location/Qualifiers  
 1..617

/organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

BASE COUNT 132 a 200 c 156 g 129 t

## ORIGIN

Query Match 84.8%; Score 511.4; DB 6; Length 617;  
 Best Local Similarity 92.1%; Pred. No. 3.9e-112;  
 Matches 550; Conservative 0; Mismatches 46; Indels 1; Gaps 1;

QY 1 CAGGAACAGCTTCTCTGCTCTCTGCACTGACAACTCACTCTGCAAGATGTC 60

Db 3 CAGGAACAGCTTCTCTGCTCTCTGCACTGACAACTCACTCTGCAAGATGTC 62

QY 61 CTGCGAGAGAACAGAGAGAGTGCACACCCCAAGTCTCCCTGCAAGATGTC 120

Db 63 CTGCGAGAGAACAGAGAGAGTGCACACCCCAAGTCTCCCTGCAAGATGTC 122

QY 121 CCAAGAGAGCCAGTACAGTGTGCTGCAAGTCTCTGCTGCTGCTGCTGCTGCTG 180



TITLE  
JOURNAL  
COMMENT

Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, T., Tirelli, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.T., Zimmer, A., and Zody, M.  
Direct Submission

Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced g1:6957762.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

Project Information

Center project name: L6490

Center clone name: 542\_E\_22

\* NOTE: This record contains 85 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
784 883: contig of 783 bp in length  
884 1703: contig of 820 bp in length  
1704 2625: contig of 822 bp in length  
2626 2725: gap of 100 bp  
2726 3512: contig of 787 bp in length  
3513 4414: contig of 802 bp in length  
4415 4514: gap of 100 bp  
4515 5306: contig of 792 bp in length  
5307 5406: gap of 100 bp  
5407 6195: contig of 789 bp in length  
6196 6295: gap of 100 bp  
6296 7099: contig of 804 bp in length  
7100 7199: gap of 100 bp  
7200 7998: contig of 799 bp in length  
7999 8098: gap of 100 bp  
8099 8891: contig of 793 bp in length  
8892 8991: gap of 100 bp  
8992 9783: contig of 792 bp in length  
9784 9883: gap of 100 bp  
9884 10701: contig of 818 bp in length  
10702 10801: gap of 100 bp  
10802 11626: contig of 825 bp in length  
11627 11726: gap of 100 bp  
11727 12530: contig of 824 bp in length  
12531 12650: gap of 100 bp  
12651 13429: contig of 779 bp in length  
13430 13529: gap of 100 bp  
13530 14321: contig of 792 bp in length  
14322 14421: gap of 100 bp  
14422 15208: contig of 787 bp in length  
15209 15308: gap of 100 bp  
15309 16072: contig of 764 bp in length  
16073 16172: gap of 100 bp  
16173 16910: contig of 738 bp in length  
16911 17010: gap of 100 bp  
17011 17815: contig of 805 bp in length  
17816 17915: gap of 100 bp  
17916 18725: contig of 810 bp in length

18726 18825: gap of 100 bp  
18826 19661: contig of 836 bp in length  
19662 19761: gap of 100 bp  
19762 20528: contig of 767 bp in length  
20529 20628: gap of 100 bp  
20629 21438: contig of 810 bp in length  
21439 21539: gap of 100 bp  
21539 22365: contig of 827 bp in length  
22366 22465: gap of 100 bp  
22466 23287: contig of 822 bp in length  
23288 23387: gap of 100 bp  
23388 24152: contig of 765 bp in length  
24153 24252: gap of 100 bp  
24253 25060: contig of 808 bp in length  
25061 25161: gap of 100 bp  
25161 25945: contig of 785 bp in length  
25946 26045: gap of 100 bp  
26046 26827: contig of 782 bp in length  
26828 26927: gap of 100 bp  
26928 27721: contig of 794 bp in length  
27722 27821: gap of 100 bp  
27822 28580: contig of 859 bp in length  
28581 28780: gap of 100 bp  
28781 29619: contig of 839 bp in length  
29620 29719: gap of 100 bp  
29720 30536: contig of 817 bp in length  
30537 30636: gap of 100 bp  
30637 31423: contig of 787 bp in length  
31424 31523: gap of 100 bp  
31524 32324: contig of 801 bp in length  
32325 32424: gap of 100 bp  
32425 33205: contig of 780 bp in length  
33206 33304: gap of 100 bp  
33305 34067: contig of 763 bp in length  
34068 34167: gap of 100 bp  
34168 34884: contig of 817 bp in length  
34885 35084: gap of 100 bp  
35085 35869: contig of 785 bp in length  
35870 36759: contig of 790 bp in length  
36760 37659: gap of 100 bp  
37660 37759: contig of 800 bp in length  
37760 38580: gap of 100 bp  
38581 38680: contig of 821 bp in length  
38681 39497: gap of 100 bp  
39498 39597: contig of 817 bp in length  
39599 40384: contig of 787 bp in length  
40385 40484: gap of 100 bp  
40485 41285: contig of 801 bp in length  
41286 41385: gap of 100 bp  
41386 42178: contig of 793 bp in length  
42179 42278: gap of 100 bp  
42279 43062: contig of 784 bp in length  
43063 43162: gap of 100 bp  
43163 43944: contig of 782 bp in length  
43945 44044: gap of 100 bp  
44045 44627: contig of 783 bp in length  
44628 44927: gap of 100 bp  
44928 45728: contig of 801 bp in length  
45729 45828: gap of 100 bp  
45829 46626: contig of 798 bp in length  
46627 46726: gap of 100 bp  
46727 47507: contig of 781 bp in length  
47508 47607: gap of 100 bp  
47608 48432: contig of 825 bp in length  
48433 48532: gap of 100 bp  
48533 49357: contig of 825 bp in length  
49358 49457: gap of 100 bp  
49458 50372: contig of 815 bp in length  
50373 50372: gap of 100 bp  
50373 51159: contig of 787 bp in length  
51160 51159: gap of 100 bp

51260	52046:	contig of 787 bp	in length
52047	52146:	gap of 100 bp	
52147	52918:	contig of 772 bp	in length
52819	53018:	gap of 100 bp	
53019	53827:	contig of 809 bp	in length
53828	53927:	gap of 100 bp	
53928	54700:	contig of 773 bp	in length
54701	54800:	gap of 100 bp	
54801	55587:	contig of 787 bp	in length
55588	55687:	gap of 100 bp	
55688	56511:	contig of 824 bp	in length
56512	57397:	contig of 786 bp	in length
57398	57497:	gap of 100 bp	
57498	58314:	contig of 817 bp	in length
58315	58414:	gap of 100 bp	
58415	59251:	contig of 837 bp	in length
59252	59351:	gap of 100 bp	
59352	60133:	contig of 782 bp	in length
60134	60237:	gap of 100 bp	
60234	61033:	contig of 800 bp	in length
61034	61133:	gap of 100 bp	
61134	61933:	contig of 799 bp	in length
61933	62032:	gap of 100 bp	
62033	62819:	contig of 787 bp	in length
62820	62919:	gap of 100 bp	
62920	63707:	contig of 788 bp	in length

[illegible][illegible]

**TITLE** Direct Submission  
**JOURNAL** Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Jul 13, 2000 this sequence version replaced gi:5957762.

----- Genome Center  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whithead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information -----  
Center project name: L6490  
Center clone name: 542\_E\_22

NOTE: This record contains 85 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allow overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

1	783:	contig of 783 bp	in length
784		gap of 100 bp	
884	1703:	contig of 820 bp	in length
1704		gap of 100 bp	
1804	2823:	contig of 822 bp	in length
2826		gap of 100 bp	
2726	3512:	contig of 787 bp	in length
3513		gap of 100 bp	
3613	4612:	contig of 802 bp	in length
4415		gap of 100 bp	
4515	5306:	contig of 792 bp	in length
5307		gap of 100 bp	

*	5407	6135:	contig of 769 bp in length
*	6196	6295:	gap of 100 bp
*	6296	7039:	contig of 804 bp in length
*	7100	7199:	gap of 100 bp
*	7200	7998:	contig of 799 bp in length
*	7999	8098:	gap of 100 bp
*	8099	8891:	contig of 793 bp in length
*	8892	8991:	gap of 100 bp
*	8992	9783:	contig of 792 bp in length
*	9784	9883:	gap of 100 bp
*	9884	10701:	contig of 818 bp in length
*	10702	10801:	gap of 100 bp
*	10802	11656:	contig of 825 bp in length
*	11627	11726:	gap of 100 bp
*	11127	12550:	contig of 824 bp in length
*	12551	12650:	gap of 100 bp
*	12651	13459:	contig of 779 bp in length
*	13430	13529:	gap of 100 bp
*	13530	14321:	contig of 792 bp in length
*	14422	14441:	gap of 100 bp
*	14422	15208:	contig of 787 bp in length
*	15409	15308:	gap of 100 bp
*	15309	16072:	contig of 764 bp in length
*	16073	16172:	gap of 100 bp
*	16173	16910:	contig of 738 bp in length
*	16911	17010:	gap of 100 bp
*	17011	17815:	contig of 805 bp in length
*	17816	17915:	gap of 100 bp
*	17916	18755:	contig of 810 bp in length
*	18726	18825:	gap of 100 bp
*	18826	19661:	contig of 836 bp in length
*	19662	19761:	gap of 100 bp
*	19762	20558:	contig of 767 bp in length
*	20529	20658:	gap of 100 bp
*	20629	21438:	contig of 810 bp in length
*	21439	21538:	gap of 100 bp
*	21539	22365:	contig of 827 bp in length
*	22366	22465:	gap of 100 bp
*	22466	23287:	contig of 822 bp in length
*	23288	23387:	gap of 100 bp
*	23388	24152:	contig of 765 bp in length
*	24153	24252:	gap of 100 bp
*	24253	25050:	contig of 808 bp in length
*	25061	25160:	gap of 100 bp
*	25161	25945:	contig of 785 bp in length
*	25946	26045:	gap of 100 bp
*	26046	26897:	contig of 782 bp in length
*	26828	26927:	gap of 100 bp
*	26828	27711:	contig of 794 bp in length
*	27722	27821:	gap of 100 bp
*	27822	28680:	contig of 859 bp in length
*	28681	28760:	gap of 100 bp
*	28781	29619:	contig of 839 bp in length
*	29620	29719:	gap of 100 bp
*	29720	30356:	contig of 817 bp in length
*	30337	30636:	gap of 100 bp
*	30637	31433:	contig of 787 bp in length
*	31424	31523:	gap of 100 bp
*	31524	32344:	contig of 801 bp in length
*	32325	32444:	gap of 100 bp
*	32425	33244:	contig of 780 bp in length
*	33205	33304:	gap of 100 bp
*	33305	34067:	contig of 763 bp in length
*	34068	34167:	gap of 100 bp
*	34168	34984:	contig of 817 bp in length
*	34985	35084:	gap of 100 bp
*	35085	35869:	contig of 785 bp in length
*	35870	35959:	gap of 100 bp
*	35970	36759:	contig of 790 bp in length
*	36760	36859:	gap of 100 bp
*	36860	37659:	contig of 800 bp in length
*	37660	37759:	gap of 100 bp
*	37760	38580:	contig of 821 bp in length

*	38581	38680:	gap of 100 bp
*	38681	39497:	contig of 817 bp in length
*	39498	39597:	gap of 100 bp
*	39598	40384:	contig of 787 bp in length
*	40385	40484:	gap of 100 bp
*	40485	41285:	contig of 801 bp in length
*	41286	41385:	gap of 100 bp
*	41386	42178:	contig of 793 bp in length
*	42179	42278:	gap of 100 bp
*	43063	43062:	contig of 784 bp in length
*	43163	43162:	gap of 100 bp
*	43945	43944:	contig of 782 bp in length
*	44045	44044:	gap of 100 bp
*	44045	44827:	contig of 783 bp in length
*	44828	44927:	gap of 100 bp
*	44928	45728:	contig of 801 bp in length
*	45729	45828:	gap of 100 bp
*	45829	46626:	contig of 798 bp in length
*	46627	46726:	gap of 100 bp
*	46727	47507:	contig of 781 bp in length
*	47508	47607:	gap of 100 bp
*	47608	48432:	contig of 825 bp in length
*	48433	48532:	gap of 100 bp
*	48533	49357:	contig of 825 bp in length
*	49358	49457:	gap of 100 bp
*	49458	50272:	contig of 815 bp in length
*	50273	50372:	gap of 100 bp
*	50373	51159:	contig of 787 bp in length
*	51160	51259:	gap of 100 bp
*	51260	52046:	contig of 787 bp in length
*	52047	52146:	gap of 100 bp
*	52147	52918:	contig of 772 bp in length
*	52919	53018:	gap of 100 bp
*	53019	53827:	contig of 809 bp in length
*	53828	53927:	gap of 100 bp
*	53928	54700:	contig of 773 bp in length
*	54701	54800:	gap of 100 bp
*	54801	55387:	contig of 787 bp in length
*	55388	55687:	gap of 100 bp
*	55688	56511:	contig of 824 bp in length
*	56512	56611:	gap of 100 bp
*	56612	57397:	contig of 786 bp in length
*	57398	57497:	gap of 100 bp
*	57498	58314:	contig of 817 bp in length
*	58315	58414:	gap of 100 bp
*	58415	59251:	contig of 837 bp in length
*	59252	59351:	gap of 100 bp
*	59352	60133:	contig of 782 bp in length
*	60134	60233:	gap of 100 bp
*	60234	61033:	contig of 800 bp in length
*	61034	61133:	gap of 100 bp
*	61134	61932:	contig of 799 bp in length
*	61933	62032:	gap of 100 bp
*	62033	62819:	contig of 787 bp in length
*	62820	62919:	gap of 100 bp
*	62920	63707:	contig of 788 bp in length

Query Match	77.2%;	Score 465.4;	DB 2;	length 76369;
Best Local Similarity	90.2%;	Pred. No. 2.6e-101;		
Matches 518; Conservative	0;	Mismatches 54;	Indels 2;	Gaps 2;

QY	19	GCCGCCCTGCACCTGAGAACTCAACTCTCTGCAGAGATGCTCTGCAGACAAACACAGA	78
Db	28106	GCTTTGTTTGTTCAGAGACAACTCAACTCTCTGCAGAGATGCTCTGCAGACAAACACAGA	28155
QY	79	GCAAGTGCACACCCCAACCCCAAGTGTCTCTACACCAAGTGTCCCCCAAGAGCCCACTACA	138
Db	28166	GCAAGTGCACACCCCAACCCCAAGTGTCTCTACACCAAGTGTCCCCCAAGAGCCCACTACA	28225
QY	139	GATGTCTGCCTCAAGCTTCTCTTGAGCTGAGCCCAAGCTCTGGAGATCTGTGAGCCCTAGCTC	198
Db	28226	GATGTCTGCCTCAAGCTTCTCTTGAGCTGAGCCCAAGCTCTGGAGAGCTGTGAGCCCTAGCTC	28285

QY 199 CGAGGCGGCTGCTTCCGTAACCAACAGGCGCCACCGAGTCCCGGCGCAGAGGCC 258  
 Db 28286 CGAGGCGGCTGCTTCCGTAACCAACAGGCGCCACCGAGTCCCGGCGCAGAGGCC 28345  
 QY 259 CAACCTCTGTGACAGGCGCAGTGTGACAGAGCGGCGGCTGTGCTGTGCTGCGCAGGTTG 318  
 Db 28346 CAACCTCTGTGACAGGCGCAGTGTGACAGAGCGGCGGCTGTGCTGTGCGGCGCAGGTTG 28405  
 QY 319 TGGGGGGTGTGCTGTGATCCAGATCTGTGATGTGAGAACAGCATCTTTGAGAGAAACAG 378  
 Db 28406 TGGGGGGTGTGCTGTGATCCAGATCTGTGATGTGAGAACAGCATCTTTGAGAGAAACAG 28464  
 QY 379 AATCCCAAGGCGCCAGAAACAGCCCATCTGACGATGCTCCATATACCTCTTTCG 438  
 Db 28465 AATCCCAAGGCGCCAGAAACAGCCCATCTGACGATGCTCCATATACCTCTTTCG 28524  
 QY 439 ACTTTCACAGGCTGAGCTGAGGTTTCTGTGGGGGATGTGAGCTCTCCCAAGAGCA 498  
 Db 28525 GCTTTCACAGGCTGAGCTGAGGTTTCTGTGGGGGATGTGAGCTCTCCCAAGAGCA 28583  
 QY 499 CTCTCTGTTTATGTACAGATGTCAATGTCCTCCATCCCTGTCCTGTCAGAGATG 558  
 Db 28584 CTCTCTGTTTATGTACAGATGTCAATGTCCTCCATCCCTGTCCTGTCAGAGATG 28643  
 QY 559 GCACTGCTTGTGCGCAACCTGTAAAGATTA 592  
 Db 28644 GCACTGCTTGTGCGCAACCTGTAAAGATTA 28677

RESULT 6  
 LOCUS AL136426  
 DEFINITION Human DNA sequence from clone Rpl-11013 on chromosome 1q21.1-21.3, complete sequence.  
 ACCESSION AL136426  
 VERSION AL136426.11 GI:16972880  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 52355)  
 Johnson, C.  
 Direct Submission  
 Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 On Nov 16, 2001 this sequence version replaced gi:16030120.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL, Sw: SWISSPROT, Tr: TrEMBL, Wp: WormPeP, Information on the WormPeP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
 This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/BSG/Chrl  
 Rpl-11013 is from the library RPl-1 constructed by the group of Pletier de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm

## COMMENT

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

VECTOR: pCYPAC2  
 IMPORTANT: This sequence is not the entire insert of clone RPl-11013. It may be shorter because we sequence overlapping sections only once, except for a short overlap.  
 The true left end of clone RPl-5210 is at 50356 in this sequence. The true right end of clone RPl-9105 is at 2000 in this sequence.  
 location/Qualifiers

## FEATURES

source  
 1. 52355  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="q21.1-21.3"  
 /clone="RPl-11013"  
 /clone\_id="RPl-1"  
 BASE COUNT 15094 a 10374 c 10652 g 16235 t  
 ORIGIN

Query Match 59.8%; Score 360.4; DB 9; Length 52355;  
 Best Local Similarity 80.8%; Pred. No. 4.4e-76;  
 Matches 459; Conservative 0; Mismatches 101; Indels 8; Gaps 3;

QY 40 CTCACTCTCTGCAAGATGTCTGTCGACAGACAGACAGAGTGCACACCCCAACCA 99  
 Db 50558 CTGATTCCTGCAAGATGTCTGTCGACAGACAGACAGAGTGCACACCCCAACCA 50617  
 QY 100 GTGTCCCTCAACCAAGTGTCTGTCGACAGACAGAGTGCACACCCCAACCA 159  
 Db 50618 GTGTCCCTCAACCAAGTGTCTGTCGACAGACAGAGTGCACACCCCAACCA 50677  
 QY 160 TGGCTGTGCCCCCAAGCTGTGAGGATCTGTGAGCTGTGAGCTGTGAGCTGTGAG 219  
 Db 50678 TGGCTGTGCCCCCAAGCTGTGAGGATCTGTGAGCTGTGAGCTGTGAGCTGTGAG 50737  
 QY 220 CCACCAAGGCTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAG 273  
 Db 50738 CCACCAAGGCTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAG 50797  
 QY 274 GGGCAGAGTGTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGAT 333  
 Db 50798 GGGCAGAGTGTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGAT 50857  
 QY 334 ATCCAGATCTGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAG 393  
 Db 50858 ATCCAGATCTGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAG 50917  
 QY 394 AGAAGAGCCCATCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGAT 452  
 Db 50918 AGAAGAGCCCATCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGAT 50977  
 QY 453 AGCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGAT 512  
 Db 50978 AGCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGAT 51037  
 QY 513 TACGATGTCTGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGAT 572  
 Db 51038 TACGATGTCTGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGATGTCTGAGGAT 51096  
 QY 573 CAACCTGTAAAGATTAAGTTCCGT 600  
 Db 51097 CAACCTGTAAAGATTAAGTTCCGT 51124

RESULT 7  
 LOCUS AL139247  
 DEFINITION Human DNA sequence from clone Rpl-5210 on chromosome 1, complete sequence.  
 ACCESSION AL139247  
 VERSION AL139247.12 GI:16444666  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 117084)  
Thomas, D

## COMMENT

Direct Submission  
Submitted (24-Oct-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequests@sanger.ac.uk  
On Oct 25, 2001 this sequence version replaced gi:15021286.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em., EMBL; Sw.,  
SWISSPROT; Tr., TrEMBL; Wp., WormPeP; Information on the WormPeP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chrl  
RPI-52J10 is from the library RPI-1 constructed by the group of  
Pleier de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pCYPAC2  
This sequence is the entire insert of clone RPI-52J10 The true left  
end of clone RPI-43017 is at 61046 in this sequence. The true right  
end of clone RPI-11013 is at 14156 in this sequence.

## FEATURES

## source

1..117084

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="1"

/clone\_lib="RPI-52J10"

/clone\_id="RPI-1"

/note="Sequence from uni-directional dGTP big dye

terminator reads only."

/note="Sequence from uni-directional dGTP big dye

terminator reads only."

/note="Sequence from uni-directional dGTP big dye

terminator reads only."

/note="Sequence from uni-directional dGTP big dye

terminator reads only."

/note="Sequence from uni-directional dGTP big dye

terminator reads only."

/note="Sequence from uni-directional dGTP big dye

terminator reads only."

/note="Sequence from uni-directional dGTP big dye

terminator reads only."

/note="Sequence from uni-directional dGTP big dye

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

## misc\_feature

69467..69488

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."

terminator reads only."



Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P.,  
 Meador, M., Mei, G., Mescher, S., Metzger, M., Miller, A., Miner, G.,  
 Minner, Z., Mitchell, T., Monabdat, K., Montgomery, K.T., Morgan, M.,  
 Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokoko, S.,  
 Oguh, M., Okunola, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
 Peery, J., Perez, L., Peters, L., Pickens, R., Pirmas, E., Puli, L.,  
 Quiler, M., Ren, Y., Rivas, M., Rojas, A., Rojiboken, I., Rolfe, M.,  
 Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shim, C.,  
 Shostet, N., Sison, I., Sodergren, E., Sonalke, T., Sparks, A.,  
 Stanley, H., Stone, H., Sutton, A., Swalek, A., Taber, P., Tameria, A.,  
 Tameria, K., Tang, H., Tansley, J., Taylor, C., Taylor, T., Telford, B.,  
 Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D.,  
 Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R.,  
 Washington, C., Wallington, S., Williams, G., Williamson, A.,  
 Wlecyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J.,  
 Zorrilla, S., Kuchelapatti, R., Weinstein, G., and Gibbs, R.  
 Direct Submission  
 Unpublished  
 3 (bases 1 to 228396)  
 Morley, K.C.  
 Direct Submission  
 Submitted (26-FEB-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 4 (bases 1 to 228396)  
 Morley, K.C.  
 Direct Submission  
 Submitted (09-APR-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Apr 9, 2003 this sequence version replaced gi:29568000.  
 -----  
 Center: Baylor College of Medicine  
 Genome Center  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 Project Information  
 Center project name: LAAY  
 Center clone name: CH250-288A21  
 -----  
 Summary Statistics  
 Sequencing vector: plasmid  
 Chemistry: Dye-terminator Big Dye, 100% of reads  
 Consensus quality: 9951 bases at least Q40  
 Consensus quality: 12209 bases at least Q30  
 Consensus quality: 14605 bases at least Q20  
 Estimated insert size: 99169; sum-of-coverage estimation  
 Quality coverage: 1x in Q20 bases; sum-of-coverage estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)  
 \* NOTE: The contigs are based on the application  
 \* of the PGI method using the Human genome (NCBI build 31)  
 \* as the comparative genome.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 228396: contig of 228396 bp in length.  
 Location/Qualifiers  
 1..228396  
 /organism="Macaca mulatta"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9544"  
 /clone="CH250-288A21"  
 1..228396  
 /notes="assembly\_name:CH250-288A21.1A  
 CONFIDENCE:\_0.83"  
 misc\_feature

BASE COUNT 5443 a 4300 c 4314 g 5247 t 209092 others  
 ORIGIN  
 Query Match 32.8%; Score 198; DB 2; Length 228396;  
 Best Local Similarity 88.0%; Pred. No. 3,76-37;  
 Matches 227; Conservative 0; Mismatches 30; Indels 1; Gaps 1;  
 -----  
 QY 341 TCTGATCTGAGACACGACGATCTTTGGAGGAAACAGATCCCAAGGCGCAAGAACAG 400  
 DB 153918 TCTTATATCGAAGATATACGATCTTTGGAGGAAACAGATCCCAAGGCGCAAGAACAG 153859  
 QY 401 CCCCAT-CTGACGATGCTTCCCATATACCTCTTGTGATTTACAGGCTGAGCTGGA 459  
 DB 153858 CCCCATCTGACGATGCTTCCCATATACCTCTTGTGATTTACAGGCTGAGCTGAG 153799  
 QY 460 GATTTCCTGTGGGGAGATCTAGCTCTCCCAAGAGGACATCTTGTATTTATGTACAGGA 519  
 DB 153798 GATTTCCTGTGGGGAGATCTAGCTCTCCCAAGAGGACATCTTGTATTTATGTACAGGA 153739  
 QY 520 TGTATATGTCCTCCCTACCTCTTACCTGACGATTTGGAGGCTTTGCTCCCACTC 579  
 DB 153738 TGTATATGTCCTCTTCCCTGTACCTGCCAAGATTTGGAGGCTTTGCTCCCACTC 153679  
 QY 580 GTAAAAAAGATTAAGTTC 597  
 DB 153678 GTCAAAAATAAAGTTTC 153661  
 -----  
 RESULT 9  
 AC132274 199220 bp DNA linear HTG 03-SEP-2002  
 LOCUS AC132274/c  
 DEFINITION Mus musculus chromosome UMR clone R24-341121, WORKING DRAFT  
 SEQUENCE 7 unordered pieces.  
 AC132274.1 GI:22657732  
 HTG: HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 -----  
 REFERENCE  
 AUTHORS McPherson, J.D. and Waterston, R.H.  
 TITLE The sequence of Mus musculus clone  
 JOURNAL Unpublished  
 2 (bases 1 to 199220)  
 McPherson, J.D. and Waterston, R.H.  
 REFERENCE  
 AUTHORS Direct Submission  
 TITLE Submitted (03-SEP-2002) Genome Sequencing Center, 4444 Forest Park  
 JOURNAL Parkway, St. Louis, MO 63108, USA  
 -----  
 COMMENT  
 -----  
 Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Contact: submissions@wustl.wustl.edu  
 Project Information  
 Center project name: W.BB0341121  
 -----  
 Summary Statistics  
 Sequencing vector: M13; 0%  
 Chemistry: Dye-terminator Big Dye, 100% of reads  
 Chemistry: Dye-terminator Big Dye, 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 197111 bases at least Q40  
 Consensus quality: 197384 bases at least Q30  
 Consensus quality: 197515 bases at least Q20  
 Insert size: 18300; agarose-gel  
 Insert size: 198257; sum-of-coverage  
 Quality coverage: 17.62 in Q20 bases; agarose-gel  
 Quality coverage: 13.20 in Q20 bases; sum-of-coverage  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently

\* consists of 7 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 231: contig of 231 bp in length  
 \* 232 331: gap of unknown length  
 \* 332 463: contig of 132 bp in length  
 \* 464 563: gap of unknown length  
 \* 564 6869: contig of 6306 bp in length  
 \* 6870 18192: contig of 11223 bp in length  
 \* 18193 42428: gap of unknown length  
 \* 42429 42528: contig of 24136 bp in length  
 \* 42529 78603: gap of unknown length  
 \* 78604 199220: contig of 36075 bp in length  
 \* 78704 199220: contig of 120517 bp in length.

## FEATURES

source

1. 199220  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="X"  
 /clone="RP24-341121"

misc\_feature  
 /note="assembly\_name:Contig29"  
 332..463

misc\_feature  
 /note="assembly\_name:Contig10"  
 564..6869

misc\_feature  
 /note="assembly\_name:Contig43"  
 6970..18192

misc\_feature  
 /note="assembly\_name:Contig44"  
 18293..42428

misc\_feature  
 /note="assembly\_name:Contig45"  
 42529..78603

misc\_feature  
 /note="assembly\_name:Contig46"  
 78704..199220

BASE COUNT 58541 a 40196 c 40401 g 59482 t 600 others

ORIGIN

Query Match 31.1%; Score 187.6; DB 2; Length 199220;  
 Best Local Similarity 76.6%; Pred. No. 1.2e-34;  
 Matches 288; Conservative 0; Mismatches 69; Indels 19; Gaps 4;

42 CAACCTCCGCAAGATGCTGCTGAGAGAAACAGAGAGAGTGCACACCCCAAGT 101

49729 CAGTTCTGCGGAGATGCTGCTGAGAGAAACAGAGAGAGTGCACACCCCAAGT 48670

102 GTCC-----CTACCCCAAGTGTCTCCCAAGAGCCCAAGTACAGTGTCTGCTC 149

49669 GCCCTTCCCAAGTGTCTCCCAAGAGTCCCAAGAGCAACAACAGTGTCTGCTC 48610

150 CAGCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 206

49693 CAGCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 49550

207 GCTGCTTCCTGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 253

49549 GCTGCTGCTGAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 48490

264 CCTTGACAG 323

49489 CTTTGACAG 49430

324 GCTGCTGCTGATCCAGATCTGATGATGATGATGATGATGATGATGATGATGATGAT 383

49429 GCTGCTGCTGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCC 49371

384 CAAGAGGCAAGACA 399

Db 49370 CAGGAGCCAGAGAAA 49355

RESULT 10  
 BD181458  
 LOCUS 275 bp DNA linear PAT 15-MAY-2003

DEFINITION  
 A gene of which expression changes in psoriasis and a method for e

ACCESSION  
 BD181458  
 BD181458.1 GI:30792376

VERSION  
 JP 2002330770-A/6.

KEYWORDS  
 Homo sapiens (human)

SOURCE  
 Homo sapiens

ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1 (bases 1 to 275)  
 Imai,Y., Wakimoto,K., Yamada,E., Chiba,H. and Okubo,K.  
 A gene of which expression changes in psoriasis and a method for e

AUTHORS  
 TITLE  
 JOURNAL  
 TANKABE, SRIYAKU CO LTD

COMMENT  
 OS Homo sapiens (human)  
 PN JP 2002330770-A/6  
 PD 19-NOV-2002  
 PF 25-MAY-2001 JP 2001156529  
 PI YUJI IMAI, KOJI WAKIMOTO, ERIKO YAMADA, HIROAKI CHIBA, KOSAKU PI

OKUBO  
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21 PC  
 C12N5/10, C12Q1/68,  
 PC G01N33/53, G01N33/56, C12N15/00, C12N5/00 CC A gene  
 of which expression changes in psoriasis and a method CC  
 for e

CC xamination directed to said gene  
 FH Key Location/Qualifiers  
 FT source 1..275  
 Location/Qualifiers

1..275  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

BASE COUNT 67 a 74 c 63 g 71 t

ORIGIN

Query Match 30.9%; Score 186.2; DB 6; Length 275;  
 Best Local Similarity 84.9%; Pred. No. 5e-34;  
 Matches 220; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

339 GATCTGTGCTGAG 398

1 GATCTGTGCTGAG 59

399 AGCCCATCTGAG 458

60 ACCCATCTGAG 119

459 AGCTTCTGCTGAG 518

120 GGGTTTCTGCTGAG 179

519 ATGCTATGTCCTGAG 578

180 ATGCTATGTCCTGAG 239

579 CGTAAAAAAGATTAAGTTC 597

240 CATCAAAAAATAAGCTCC 258

RESULT 11  
 AC137251 195503 bp DNA linear HTG 19-NOV-2002

LOCUS AC137251

AC137251

DEFINITION	Rattus norvegicus clone CH230-unknown, *** SEQUENCING IN PROGRESS
ACCESSION	AC137251
VERSION	GI:25073159
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 195503)
AUTHORS	Munry,D,Marle, Metzker,M,Lee, Abramzon,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooks,S, Amit,A, Anguiano,D, Aryalbechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Behnmed,F, Biwalto,K, Blair,C, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Chen,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K, Draper,H, Dugar-Rocha,S, Dunn,A, Durbin,K, Duval,B, Evans,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Georgeogis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W, Gunaratne,P, Haaland,W, Hamil,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Huily,S, Hume,J, Idlibird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levay,J, Lewis,L, Li,Z, Liu,J, Liu,T, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenshuwa,L, Loulsged,H, Lozdo,R,T, Lu,X, Ma,J, Meheswari,M, Mahindaratne,M, Mahmoud,K, Malloy,K, Mangum,A, Mangun,B, Mapa,P, Martin,K, Martin,K, Martinez,E, Mamwhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Mijda,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Munidas,M, Murphy,M, Neir,L, Nervens,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nsoakemeleh,O, Okumu,G, Olarnunagoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perre,A, Perez,L, Pfannkuch,C, Plopper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L,L, Puzo,M, Quiroz,U, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S, Sanders,W, Savery,G, Scheerer,S, Scott,G, Shatman,S, Shetty,J, Shvartsbeyn,A, Sison,I, Sitter,C,D, Snajs,D, Sneed,A, Sodergren,E, Song,X,-Z, Sorelle,R, Sosa,J, Steinle,M, Strong,R, Sutton,A, Svatek,A, Taboc,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K, Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willson,R, Wlaczek,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von Niederhausen,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinsteck,G, and Gibbs,R,A.
TITLE	Unpublished
JOURNAL	2 (bases 1 to 195503)
REFERENCE	Rat Genome Sequencing Consortium.
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL	The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas ( <a href="http://www.hgsc.bcm.tmc.edu/projects/rat/">http://www.hgsc.bcm.tmc.edu/projects/rat/</a> ). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,
COMMENT	

Individual sequence contigs are ordered and oriented, and separately sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- genome center -----  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information -----  
Center project name: KZIE  
Center clone name: CH230-unknown

```
----- Summary Statistics -----
Assembly program: Phrap; version 0.990339
Consensus quality: 164059 bases at least Q40
Consensus quality: 160085 bases at least Q30
Consensus quality: 167609 bases at least Q20
Estimated insert size: 171122; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
```

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.igsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.igsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

180274	180273: contig of 180273 bp in length
180374	180373: gap of unknown length
181900	181899: contig of 1526 bp in length
181999	181999: gap of unknown length
182000	195503: contig of 13504 bp in length.

```

FEATURES
    source
        1..195503
            location=Qualifiers
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-unknown"
            1..1873
                misc_feature
                    /note="Wgs contig"
                    174826..176310
                        misc_feature
                            /note="Wgs contig"
BASE COUNT  50392 a 34104 c 34657 g 48942 t 27408 others
ORIGIN

```

30.7%; Score 185.4; DB 2; Length 195503;  
Similarity 77.0%; Pred. No. 3,9e-34;  
0; Mismatches 66; Indels 19; Gaps 4;

6 TCCTCCCAAGATGTCTCTGCACAGACGACGACGTGCCAACCCCAACCAAGTGTCC 105  
9 TCCTCCCAAGATGTCTCTGCACAGACGACGACGACGTGCCAACCCCAACCAAGTGTCC 159  
CTCA-----CCCAAGTGTCCCAAGACCCCAAGACCCCAAGTGTCCGCTCGAGC 153

9 CTCCTCCGAGTCTCTCCCAAGTCCCCCAAGACACAGCACAGTGTCTGCTGCAGC 1603  
4 TTCTCTGCTGTGCCCCAAGCTCTGGGGTCT---GTGGCCTTAGCTCCGAGGGCGGCTG 210

1 CTTCTGAACCAACAAGGCG--CAACCAAGATCCGGCGCAAGAGCCCACTCTG 267  
9 CTGCTTAGACCAACAAGGCGAGGCTCCCAAGATCAAGGCGGAGAGCTCCAGTTCTG 1615  
9 CTCCTTCGATGTGTGTAACAAGTCTTGGGGGCTGCAATGTCCCAAGCTCTAAGGAGAGCT 1603

8 TGACAGGGCAATGATCAGCAAGCGGGGCTTGGCTGCTGCACAGGTTCTGGGGGCTG 327

Db 16159 TGACCGTCGAGTGTGTCAGCAGTCTGGGGGCTCAGGCTGTGTGTCAACCTCTGGGGGCTG 16218

Qy 328 CTGCTGATCCAGATCCGATGCTGAGACAAAGCGATCTTTGGAGGAAAACAATCCCAAG 387

Db 16219 CTGCTAATCTGGAACATGATCTGTAGATCAAGGGA-GTCGGGAGGAAAAAAGATCCCAAA 16277

Qy 388 AGGCCAAGA 396

Db 16278 GGTCAAGA 16286

RESULT 12

AC095371 261316 bp DNA linear HTG 13-NOV-2002

LOCUS Rattus norvegicus clone CH230-1G20, \*\*\* SEQUENCING IN PROGRESS \*\*\*

DEFINITION 6 uncloned pieces.

AC095371.4 GI:24941586

VERSION HTG, HTGS PHASE1, HTGS DRAFT, HTGS\_ENRICHED.

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 261316)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amarunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonini,D., Bouck,J., Bowe,S., Brieva,M., Brown,M., Brown,M., Bryant,N.P., Butay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davis,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Derr,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homai,P., Howard,S., Huber,J., Huiy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,V., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulsged,H., Locado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheswari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Maasey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwomog, G., Ogrunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Py L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Soederren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Swalek,A., Tabor,P., Tamerisa,X., Tang,H., Tansy,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umanu,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,C., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleceyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G., and Gibbs,R.

TITLE Unpublished

JOURNAL 2 (bases 1 to 261316)

REFERENCE Direct Submission

AUTHORS Worley,K.C.

TITLE Submitted (16-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

# REFERENCE

3 (bases 1 to 261316)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Nov 13, 2002 this sequence version replaced gi:23269262. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

## ----- Genome Center

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

## ----- Project Information

Center project name: TUIT  
Center clone name: CH230-1G20  
Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 228141 bases at least Q40  
Consensus quality: 225987 bases at least Q30  
Consensus quality: 231269 bases at least Q20  
Estimated insert size: 241147; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

	1	2	3	4	5	6
3233: contig of 3233 bp in length	3234	3333	gap of unknown length	218784	gap of 215451 bp in length	218785
218785: gap of unknown length	218785	222557	contig of 3373 bp in length	222557	gap of unknown length	222558
222558: gap of unknown length	222558	223507	contig of 1150 bp in length	223507	gap of unknown length	223508
223508: gap of unknown length	223508	224607	contig of 1000 bp in length	224607	gap of unknown length	224608
224608: gap of unknown length	224608	261316	contig of 36603 bp in length.			

## FEATURES

### source

1. 261316  
/organism="Rattus norvegicus"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-1G20"

### misc\_feature

1. 1274  
/note="wgs\_contig"

### misc\_feature

3334. 4756  
/note="wgs\_contig"

### misc\_feature

218885. 219956  
/note="wgs\_contig"

### misc\_feature

220761. 222257  
/note="wgs\_contig"

BASE COUNT 67615 a 48827 c 47477 g 67871 t 29526 others

## ORIGIN

Query Match	30.7%	Score 185.4	DB 2	Length 261316
Best Local Similarity	77.0%	Pred. No. 3.8e-34		
Matches 284	Conservative	0	Mismatches 66	Indels 19
			Gaps	4
Oy	46	TCCTGCCAGATGTCCTGCCAGAGAACACAGACAGTGCACACCCCAAGTGC	105	
Db	81900	TCCTGCCAGATGTCCTGCCAGAGAACACAGACAGTGCACACCCCAAGTGC	81955	
Oy	106	CTCA-----CCAGTGTCCCAAGGCCAGTACAGTGTGCTCCAGC	153	
Db	81960	CTCCCCGAAGTGTCTCCCAAGGTCCCAAGAACAGACAGTGTGCTCCAGC	82015	
Oy	154	TTCTCTGTGCTGTGCTCCCAAGTGTGAGTCT--GTGAGCTTGTGCTCCAGGCGAGT	210	
Db	82020	CTTCTCTGTGCTGTGCTCAAGTCTGTGAGGCGGTGACAGTGTCCCAAGTGTGAGGCGAGT	82075	
Oy	211	CTTCTGTACCAACCAACAGAGCC--CACACCAATGCGGCGCCAGAGCCCAATCTCTG	267	
Db	82080	CTGCTGTAGCCACCAACAGAGCCAGTCCACAGATGACGAGCGAGAGCTCCAGTCTCTG	82135	
Oy	268	TGAAGAGGAGCATGTGTGACAGAGGCGGAGGCTGTGCTGTGCGCACCGTCTGAGGAGCT	327	
Db	82140	TGACCGTGTGAGCATGTGTGACAGAGTGTGAGGCGGTGAGCTGTGCTGTGAGGAGCT	82195	
Oy	328	CTGCTGTATCCAGATCTGTGATCTGTGACAGACAGCATCTTTGAGAGAAACAAGATCCCAAG	387	
Db	82200	CTGCTGTATCTGTGACAGATGTGTGATCTGTGACAGAGGAGTGTGAGAGAGAGAGAGATCCCAAG	82255	
Oy	388	AGCGACAGA 396		
Db	82259	GGTCAGAGA 82267		
RESULT 13				
AC118786/c				
LOCUS				
DEFINITION	Rattus norvegicus clone CH230-364n10, WORKING DRAFT SEQUENCE, 6			
ACCESSION	AC118786	347801 bp	DNA	linear
VERSION	AC118786.9	GI:250092277		
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.			
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
REFERENCE	1 (bases 1 to 347801)			
AUTHORS	Murphy D,Marie, Metzker,M,Lee, Abramzon,S, Adams,C, Alder,J, Allen,C, Allen,H, Albrooks,S, Amin,A, Anguiano,D, Anyalabechi,V, Aoyagi,A, Ayodeji,M, Baca,Z, Baden,H, Baldwin,K, Bandazanaik,D, Barber,M, Barnstead,M, Benahmed,F, Biwalo,K, Blair,D, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,F, Burrell,K, Calderon,E, Cadena,V, Carter,K, Cavaco,I, Ceasar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dedrich,D, Delgado,O, Denon,S, Deramo,C, Ding,Y, Dinh,H, Diya,K, Drepper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,W, Gergeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,M, Gunaratne,P, Haaland,N, Hamli,C, Hamilton,C, Hamilton,K, Harvey,T, Havlik,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Hulik,S, Hume,J, Idelbrd,R, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Kapachy,S, Kelly,S, Khan,Z, King,L, Koyar,C, Kowitz,C, Kieft,C,L, Lebow,H, Levay,L, Lewis,L, Li,Z, Liu,J, Liu,Y, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenshewa,L, Louisedge,H, Lozad,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindaratne,M, Mahmou,M, Malloy,K, Mangum,A,			

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Muidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwokediameh, O., Okunolu, G., Olajunwasogun, A., Pal, S.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pl, L.,  
Pruzo, K., Quirio, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, K., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,  
Sanders, W., Saverly, G., Scherer, S., Scott, G., Shamsan, S., Shen, H.,  
Shetty, J., Shattarbeyan, A., Sisson, I., Sitter, C.D., Smays, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Soes, J.,  
Steinle, M., Strong, R., Sutton, A., Swatek, A., Taber, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usman, K.,  
Valas, R., Vora, V., Villasa, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willison, R., Wleczyk, R., Woodson, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhuesern, A., Weise, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
Weinstock, G., and Gibbs, R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 347801)  
Worley, K.C.  
Direct Submission  
Submitted (21-APR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 347801)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Nov 15, 2002 this sequence version replaced gi:23269758.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.bsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separate  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GVDP  
Center clone name: CH230-364N10  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 314830 bases at least Q40  
Consensus quality: 316638 bases at least Q30  
Consensus quality: 321617 bases at least Q20  
Estimated insert size: 328293; sum-of-contigs estimation  
Quality coverage: 7X in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 6 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 224743: contig of 224743 bp in length  
\* 224744 224843: gap of unknown length  
\* 224844 240367: contig of 115524 bp in length  
\* 240368 340467: gap of unknown length  
\* 340468 342101: contig of 1634 bp in length  
\* 342102 342201: gap of unknown length  
\* 342202 343356: contig of 1155 bp in length  
\* 343357 344984: gap of unknown length  
\* 344985 345084: gap of unknown length  
\* 345085 347801: contig of 2717 bp in length.  
Location/Qualifiers

source  
1. 347801  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-364N10"

misc\_feature  
1. 1485  
/note="wgs\_contig"

misc\_feature  
35987. 37516  
/note="wgs\_contig"

misc\_feature  
227958. 229240  
/note="wgs\_contig"

misc\_feature  
264424. 265451  
/note="wgs\_contig"

BASE COUNT 98908 a 63478 c 63910 g 96687 t 24818 others  
ORIGIN

Query Match 30.7%; Score 185.4; DB 2; Length 347801;  
Best Local Similarity 77.0%; Pred. No. 3.7e-34;

Matches 284; Conservative 0; Mismatches 66; Indels 19; Gaps 4;

QY 46 TCCTGCGAGATGCTCTGCGAGCAAGACAGACAGAGTCCACCCCAAGTGTCC 105

Db 166967 TCCTGCGAGATGCTCTGCGAGCAAGACAGACAGAGTCCACCCCAAGTGTCC 166908

QY 106 CTC-----CCCAAGTGTCCCAAGAGCCCAAGTGTGTGCTTCCAGC 153

Db 166907 CTCCTGCGAGTCTCCCAAGTGTCCCAAGAGCCCAAGTGTGTGCTTCCAGC 166848

QY 154 TTCTCTGCTGTGCTCCCAAGTGTGAGTCT--GTGACCTAGTCTCCAGGAGCGCTG 210

Db 166847 CTCTTCTGCTGTGCTCAAGCTGTGAGGAGTGTCCCAAGTGTCTTCAAGGAGCTG 166788

QY 211 CTCTCTGAGCAACCAAGGCGC--CAACACCAAGTCCGAGGCGCAAGGCGCAACTGCTG 267

Db 166787 CTGCTTGAAGCCACCAAGGCGCGCAAGTCCCAAGTGAAGGCGCAAGTCTCAAGTCTG 166728

QY 268 TGACAGGAGGAGTGTGAGCAAGGAGGAGGCTGTGCTGTGCGACGCTTCTGAGGAGCTG 327

Db 166727 TGACAGGAGGAGTGTGAGCAAGGAGGAGGCTGTGCTGTGCGACGCTTCTGAGGAGCTG 166668

QY 328 CTGCTGATCCAGATCTGATGCTGAGCAAGAGCGATCTTTGAGAGAAAGAAATCCCAAG 387

Db 166667 CTGCTGATCCAGATCTGATGCTGAGCAAGAGCGATCTTTGAGAGAGAAAGAAATCCCAAA 166609

QY 388 AGGCCAAGA 396

Db 166608 GGTCAAGA 166600

RESULT 14

AF176515 566 bp mRNA linear ROD 24-MAY-2001

LOCUS AF176515 Mus musculus E1g3 protein (E1g3) mRNA, complete cds.

ACCESSION AF176515 AF303034

VERSION AF176515.2 GI:14195705

KEYWORDS Mus musculus (house mouse)

## ORGANISM

Mus musculus

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## AUTHORS

1 (bases 1 to 566)  
Mang, A., Johnson, D.G. and MacLeod, M.C.

## TITLE

Molecular cloning and characterization of a novel mouse epidermal

## JOURNAL

Genomics 73 (3), 284-290 (2001)

## MEDLINE

21248673

## PUBMED

11350120

## AUTHORS

2 (bases 496 to 566)  
Mang, A.J., Pierce, A., Judeon-Kremer, K., Gaddis, S., Aldaz, C.M.,

## TITLE

Johnson, D.G. and MacLeod, M.C.

## JOURNAL

3 (bases 1 to 566)  
Wang, A.J., Shen, J.J. and MacLeod, M.C.

## REFERENCE

Submitted (06-SEP-2000) Carcinogenesis, MD Anderson Cancer Center,

## JOURNAL

Science Park, P.O. Box 389, Smithville, TX 78957, USA

## COMMENT

On May 24, 2001 this sequence version replaced gi:5823356.

## FEATURES

Location/Qualifiers

## source

1. 566  
/organism="Mus musculus"  
/mol\_type="mRNA"

## CDS

/strain="SIN"  
/db\_xref="taxon:10090"  
/cell\_type="keratinocyte"  
/tissue\_type="dorsal skin"

## gene

<1..566  
/gene="E1g3"

## BASE COUNT

105 a 181 c 147 g 133 t

## Query Match

30.6%; Score 184.4; DB 10; Length 566;  
Best Local Similarity 77.3%; Pred. No. 1.3e-33;

## Matches

283; Conservative 0; Mismatches 61; Indels 22; Gaps 4;

## QY

11 CTCTCTGCTCTCTCTGCACTGCACTCAACTCAACTCTCTGCAAGATGCTCTGCAAGAG 70

## Db

27 CTCTCTGCTCTCTCTGCACTGCACTCAACTCAACTCTCTGCAAGATGCTCTGCAAGAG 82

## QY

71 AACGAGGAGAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAG 118

## Db

83 ACTCAAGAGAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAG 142

## QY

119 CCCCCAAGAGAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAG 178

## Db

143 CCCCCAAGAGAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAG 202

## QY

179 GGGGCTCT--GTGGGCTCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAG 232

## Db

203 GGGGCTCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAG 262

## QY

233 CACCAAGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 292

## Db

263 TCCTCAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 322

## QY

293 GGGGCTCTGCTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAG 352

## Db

323 GGGGCTCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAG 382

## QY

353 GACCAAG 358

Db 383 GACAG 388

RESULT 15  
AL139247/c  
LOCUS  
DEFINITION  
AL139247 117084 bp DNA linear PRI 24-OCT-2001  
Human DNA sequence from clone RPI-52J10 on chromosome 1, complete  
sequence.

ACCESSION  
AL139247  
VERSION  
AL139247.12 GI:16444666  
KEYWORDS  
HTG.

SOURCE  
Homo sapiens (human)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 117084)

COMMENT  
Direct Submission  
Submitted (24-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
request: clonerequest@sanger.ac.uk  
On Oct 25, 2001 this sequence version replaced gi:15021286.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em., EMBL; Sw.,  
SWISSPROT; Tr., TrEMBL; Mp., WormPEP; Information on the WormPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr1  
RPI-52J10 is from the library RPI-1 constructed by the group of  
Pierer de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pCYPAC2

This sequence is the entire insert of clone RPI-52J10 The true left  
end of clone RPI-43017 is at 61046 in this sequence. The true right  
end of clone RPI-11013 is at 14156 in this sequence.  
Location/Qualifiers

## FEATURES

source

1..117084

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"

misc\_feature

/clone\_lib="RPI-52J10"  
/clone\_lib="RPI-1"

misc\_feature

/note="Sequence from uni-directional dGTP big dye  
terminator reads only."

/note="Sequence from uni-directional dGTP big dye  
terminator reads only."

BASE COUNT 35165 a 24224 c 24423 g 33272 t  
ORIGIN

Query Match 30.5%; Score 184; DB 9; Length 117084;  
Best Local Similarity 75.4%; Pred. No. 9e-34;  
Matches 264; Conservative 0; Mismatches 65; Indels 21; Gaps 2;

QY 31 CTGGAACAATCTAACTCTGCGCAAGATGTCTGCGACAGAACGAGAGTGCACACC 90  
Db 22615 CCAGGCTCCCTGACCTCTGCTGATGTCTGCGACAGAACGAGAGTGCACAGCC 22556  
QY 91 CCCAGCAAGTGTCCCTCACCCAAAGTGTCCCCCAAGAGCCGATAGTGTCTGCTCC 150  
Db 22555 TCCGCCAAGTCCCTG-----CAAGAGCCGACAGAGTGTCTGCTCC 22511  
QY 151 AGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 210  
Db 22510 AGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 22451  
QY 211 CTGCTGACCAACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 264  
Db 22450 CTGCTGACCAACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 22391  
QY 265 CTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 324  
Db 22390 CTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 22331  
QY 325 CTGCTGCTGATTCAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 374  
Db 22330 CTGCTGCTGATTCAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 22281

Search completed: February 16, 2004, 07:16:47  
Job time : 3265 secs